Recovering quantum gates from few average gate fidelities – Supplemental Material

I. Roth,[1] R. Kueng,[2] S. Kimmel,[3] Y.-K. Liu,[4,5] D. Gross,[6] J. Eisert,[1] and M. Kliesch[7]

[1]*Dahlem Center for Complex Quantum Systems, Freie Universität Berlin, Germany*
[2]*Institute for Quantum Information and Matter, California Institute of Technology, Pasadena, USA*
[3]*Department of Computer Science, Middlebury College, USA*
[4]*National Institute of Standards and Technology, Gaithersburg, USA*
[5]*Joint Center for Quantum Information and Computer Science (QuICS), University of Maryland, College Park, USA*
[6]*Institute for Theoretical Physics, University of Cologne, Germany*
[7]*Institute of Theoretical Physics and Astrophysics, National Quantum Information Centre, University of Gdańsk, Poland*

CONTENTS

In this Supplemental Material we provide proofs and further details of the results of the main text. Section A- C develop the prerequisites to prove the recovery guarantee, Theorem 2, in Section D. The optimality of this result is addressed in Section E. The expansion of unital maps in terms of a unitary 2-design, Proposition 1, is derived in Section F. In Section G, we show that the unitarity of a hermiticity preserving map can be expressed as the variance of its average gate fidelity with respect to a unitary 2-design. We also discuss possible implications. Finally, Section H provides further details and results of the numerical demonstration of the protocol.

We start by specifying the notation that is used subsequently. For a vector space V we denote the space of its endomorphisms by $\mathrm{L}(V)$. In particular, let H_d denote the space of hermitian operators on a d-dimensional complex Hilbert space. We label the vector space of endomorphisms on H_d by $\mathrm{L}(H_d)$ and denote its elements with calligraphic letters. For every map $\mathcal{X} \in \mathrm{L}(H_d)$, we define its adjoint $\mathcal{X}^\dagger \in \mathrm{L}(H_d)$ with respect to the Hilbert-Schmidt inner product $(\cdot,\cdot)$ on H_d. We denote the subset of *completely positive* maps by $\mathrm{CP}(H_d) \subset \mathrm{L}(H_d)$. *Quantum channels* are elements of $\mathrm{CP}(H_d)$ that are *trace preserving* (TP), i.e. $\mathrm{Tr}(\mathcal{E}(X)) = \mathrm{Tr}(X)$ for all $X \in H_d$. This condition is equivalent to the identity matrix $\mathrm{Id} \in H_d$ being a fixed point of the adjoint channel, $\mathcal{E}^\dagger(\mathrm{Id}) = \mathrm{Id}$. Similarly, a map (or channel) $\mathcal{E}$ that itself has the identity as a fixed-point, $\mathcal{E}(\mathrm{Id}) = \mathrm{Id}$, is called *unital*. The affine subspace of TP and unital maps is denoted by $\mathrm{L}_{\mathrm{u,tp}}(H_d) \subset \mathrm{L}(H_d)$. We further denote the linear hull of $\mathrm{L}_{\mathrm{u,tp}}(H_d)$ by $\mathrm{L}_{\overline{\mathrm{u,tp}}}(H_d)$.

Most of our results feature a norm on $\mathrm{L}(H_d)$, which is naturally induced on by the average gate fidelity (AGF) (1) in the following way. We define the inner product on $\mathrm{L}(H_d)$ as

$$(\mathcal{X},\mathcal{Y}) = \frac{d+1}{d} F_{\mathrm{avg}}(\mathcal{X},\mathcal{Y}) - \frac{1}{d^2}(\mathcal{X}(\mathrm{Id}),\mathcal{Y}(\mathrm{Id})) \tag{S.1}$$

and denote the induced norm on $\mathrm{L}(H_d)$ by $\|\mathcal{X}\|^2 = (\mathcal{X},\mathcal{X})$. The pre-factors are chosen such that unitary channels $\mathcal{U} \in \mathrm{L}(H_d)$ have unit norm.

Note that this inner product is proportional to the previously defined Hilbert-Schmidt inner product applied to the Choi and Liouville representations:

$$(\mathcal{X},\mathcal{Y}) = (J(\mathcal{X}), J(\mathcal{Y})) = \frac{1}{d^2}(\mathcal{L}(\mathcal{X}),\mathcal{L}(\mathcal{Y})), \tag{S.2}$$

see Refs. [1, 2] and also [3, Proposition 1]. We choose the convention that Choi matrices of quantum channels have unit trace, i.e. $\mathrm{Tr}(J(\mathcal{X})) = 1$. Furthermore, for $X \in H_d$ we will encounter the Schatten norms $\|X\|_1 = \mathrm{Tr}[\sqrt{XX^\dagger}]$, $\|X\|_2 = \sqrt{\mathrm{Tr}(XX^\dagger)}$ and $\|X\|_\infty = \sqrt{\mu_{\max}(XX^\dagger)}$, where $\mu_{\max}(Y)$ denotes the maximum eigenvalue of a Hermitian matrix Y. For a vector $y \in \mathbb{R}^m$ and $q \in \mathbb{N}$ the ℓ_q-norm is defined by $\|y\|_{\ell_q} = (\sum_{i=1}^m |y_i|^q)^{1/q}$.

For a map $\mathcal{T} : H_d \to H_d$ we define the random variable

$$S_{\mathcal{T}} = d^2(\mathcal{T},\mathcal{U}) \tag{S.3}$$

where $\mathcal{U}$ is a unitary channel $\mathcal{U}(X) = UXU^\dagger$ with U either chosen uniformly at random from the full unitary group $\mathrm{U}(d)$, or the Clifford group $\mathrm{Cl}(d)$, depending on the context. The main technical ingredients for the the proofs of our main results are an expression for the second and fourth moment of $S_{\mathcal{T}}$. To this end, an integration formula for the first four moments over the Clifford group is developed in Section A. We then derive an explicit expression for the second moment of $S_{\mathcal{T}}$ in Section B and an upper bound on the fourth moment of $S_{\mathcal{T}}$ in Section C. These bounds are essential prerequisites for applying strong techniques from low-rank matrix reconstruction to prove our recovery guarantee, Theorem 2, for unitary gates in Section D.

A. An integration formula for the Clifford group

One of the main technical ingredients of the proof is an explicit formula for integrals of the diagonal action of the Clifford group $\mathrm{Cl}(d)$. More precisely, for a unitary representation $R : G \to \mathrm{L}(V)$ of a subgroup $G \subset \mathrm{U}(d)$ carried by a vector space V, we define $E_R : \mathrm{L}(V) \to \mathrm{L}(V)$ ("twirling") as

$$E_R(A) = \int_G R(g) A R(g)^\dagger d\mu(g), \tag{S.4}$$

where μ is the invariant measure induced by the Haar measure on $\mathrm{U}(d)$.

For $V = (\mathbb{C}^d)^{\otimes n}$ we denote the diagonal action of a subgroup G of $\mathrm{GL}(\mathbb{C}^d)$ by $\Delta_G^n : G \to \mathrm{GL}(V)$, i.e.

$$\Delta_G^n : U \mapsto \underbrace{U \otimes \ldots \otimes U}_{n \text{ times}}. \tag{S.5}$$

Note that if G is a subgroup of the unitary group $\mathrm{U}(d)$ then Δ_G^n is a unitary representation. The main result of this chapter is an explicit expression for $E_{\Delta^4_{\mathrm{Cl}(d)}}(A)$ for arbitrary $A \in \mathrm{L}(V)$.

For $E_{\Delta^n_{\mathrm{U}(d)}}(A)$, where the integration is carried out over the entire unitary group, an explicit formula was derived in Refs. [4, 5]. It is instructive to review the result of Ref. [5] and its proof first. Our derivation of the analogous expression for the Clifford group follows the same strategy and makes use of many of the intermediate results.

1. *Integration over the unitary group* $\mathrm{U}(d)$

To state the result we have to introduce notions from the representation theory of $\Delta^n_{\mathrm{U}(d)}$ which can be found, e.g., in Refs. [4–7]. Schur-Weyl duality relates the irreducible representations of the diagonal action of $\mathrm{GL}(V)$ to the irreducible representations of the natural action of the symmetric group S_n on V. Recall that the representation $\Delta^n_{\mathrm{U}(d)}$ decomposes into irreducible representations $\Delta^\lambda_{\mathrm{U}(d)} : \mathrm{U}(d) \to \mathrm{GL}(W_\lambda)$ labelled by partitions $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_{l(\lambda)})$ of n into $l(\lambda) \leq d$ integers, i.e. $\sum_{i=1}^{l(\lambda)} \lambda_i = n$. For short, we denote a partition of n by $\lambda \vdash n$ and dimensions of the Weyl-modules W_λ by D_λ.

Let $\{\,|i\rangle\}_{i=1}^d$ be an orthonormal basis of $\mathbb{C}^d$. We define the representation $\pi^d_{S_n} : S_n \to \mathrm{GL}(V)$ by linearly extending

$$\pi^d_{S_n}(\tau) : |i_1\rangle \otimes \ldots \otimes |i_k\rangle \mapsto \left|i_{\tau^{-1}(1)}\right\rangle \otimes \ldots \otimes \left|i_{\tau^{-1}(k)}\right\rangle. \tag{S.6}$$

The irreducible representations of $\pi^d_{S_n}$, $\pi^\lambda_{S_n} : S_n \to \mathrm{GL}(S_\lambda)$ are also labelled by partitions $\lambda \vdash n$. The dimensions of the Specht-modules S_λ are denoted by d_λ. Since the actions of $\Delta^n_{\mathrm{U}(d)}$ and $\pi^d_{S_n}$ commute, they induce a representation of $\mathrm{U}(d) \times S_n$ on $(\mathbb{C}^d)^{\otimes n}$ that decomposes into irreducible representations as follows:

Theorem S.1 (Schur-Weyl decomposition). *The action of $\mathrm{U}(d) \times S_n$ on $(\mathbb{C}^d)^{\otimes n}$ is multiplicity free and $(\mathbb{C}^d)^{\otimes n}$ decomposes into irreducible components as*

$$(\mathbb{C}^d)^{\otimes n} \cong \bigoplus_{\lambda \vdash n, l(\lambda) \leq d} W_\lambda \otimes S_\lambda \tag{S.7}$$

on which $\mathrm{U}(d) \times S_n$ acts as $\Delta^\lambda_{\mathrm{U}(d)} \otimes \pi^\lambda_{S_n}$.

We denote the orthogonal projections on $W_\lambda \otimes S_\lambda$ by P_λ and the character on the irreducible representation $\pi^\lambda_{S_n}$ of S_n by $\chi^\lambda(\pi) := \mathrm{Tr}(\pi^\lambda_{S_n}(\pi))$. The orthogonal projectors can be written as

$$P_\lambda = \frac{d_\lambda}{n!} \sum_{\sigma \in S_n} \chi^\lambda(\sigma) \pi^d_{S_n}(\sigma), \tag{S.8}$$

see, e.g. Ref. [8, Eq. (12.10)]. In terms of these projectors $E_{\Delta^n_{\mathrm{U}(d)}}(A)$ can be calculated using the following theorem.

Theorem S.2 (Integration over the unitary group $\mathrm{U}(d)$). *Let $A \in \mathrm{L}(V)$. Then, for $R = \Delta^n_{\mathrm{U}(d)}$ and $G = \mathrm{U}(d)$,*

$$\begin{aligned} &E_{\Delta^n_{\mathrm{U}(d)}}(A) \\ &\quad = \frac{1}{n!} \sum_{\tau \in S_n} \mathrm{Tr}(A \pi^d_{S_n}(\tau))\, \pi^d_{S_n}(\tau^{-1}) \sum_{\lambda \vdash n,\, l(\lambda) \leq d} \frac{d_\lambda}{D_\lambda} P_\lambda. \end{aligned} \tag{S.9}$$

This formula differs slightly from the original statement presented in Ref. [5]. The more common formulation presented there follows from evaluating the expression of Theorem S.2 using a standard tensor basis of $\mathrm{L}(V)$ [9]. However, here we have opted for a presentation of Theorem S.2 that is easier to generalise beyond the full unitary group.

In the remainder of this section, we present a proof of Theorem S.2 following the strategy of Ref. [5]. The commutant of a subset $\mathcal{A} \subset \mathrm{L}(V)$ is the subset of $\mathrm{L}(V)$ defined by

$$\mathrm{Comm}(\mathcal{A}) = \{B \in \mathrm{L}(V) \mid BA = AB \quad \forall A \in \mathcal{A}\}. \tag{S.10}$$

It is straight-forward to verify the following well-known properties of E_R:

Lemma S.3 (Properties of E_R). *Let R be a unitary representation of a subgroup $G \subseteq \mathrm{U}(d)$. Then, for all $A \in \mathrm{L}(V)$ and $B \in \mathrm{Comm}(R(G))$, the map E_R (defined in Eq. (S.4)) fulfils*

$$\mathrm{Tr}(E_R(A)) = \mathrm{Tr}(A), \tag{S.11}$$

$$E_R(AB) = E_R(A) B, \tag{S.12}$$

$$E_R(A) \in \mathrm{Comm}(R(G)). \tag{S.13}$$

The last statement of Lemma S.3 implies that $E^n_{\Delta_{\mathrm{U}(d)}}(A)$ is in the commutant of $\Delta^n_{\mathrm{U}(d)}$ for all $A \in \mathrm{L}(V)$. Using the decomposition of Theorem S.1 and Schur's Lemma we therefore conclude that $E_{\Delta^n_{\mathrm{U}(d)}}(A)$ acts as the identity on the Weyl-modules,

$$E_{\Delta^n_{\mathrm{U}(d)}}(A) = \sum_{\lambda \vdash n, l(\lambda)\leq d} \mathrm{Id}_{D_\lambda} \otimes E_\lambda \tag{S.14}$$

with $E_\lambda \in \mathrm{L}(S_\lambda)$. In general, the direct sum of endomorphisms acting on the irreducible representations of a group is isomorphic to the group ring which consists of formal (complex) linear combinations of the group elements [7, Propositon 3.29]. We denote the group ring of S_n by $\mathbb{C}[S_n]$.

To derive an explicit expression of the coefficient of the expansion of $E_{\Delta^n_{\mathrm{U}(d)}}(A)$ in $\mathbb{C}[S_n]$, we introduce the map $\Phi : \mathrm{L}(V) \to \mathrm{L}(V)$

$$\Phi(A) = \sum_{\sigma\in S_n} \mathrm{Tr}(A\pi^d_{S_n}(\sigma^{-1}))\pi^d_{S_n}(\sigma). \tag{S.15}$$

We will make use of the following properties of the map Φ.

Lemma S.4 (Properties of Φ)**.** *For all $A \in \mathrm{L}(V)$ and $B \in \mathrm{Comm}(\Delta^n_{\mathrm{U}(d)})$*

$$\Phi(A) = \Phi(E_{\Delta^n_{\mathrm{U}(d)}}(A)), \tag{S.16}$$

$$\Phi(B) = B\Phi(\mathrm{Id}), \tag{S.17}$$

$$\Phi(\mathrm{Id})^{-1} = \frac{1}{n!}\sum_{\lambda\vdash n, l(\lambda)\leq d} \frac{d_\lambda}{D_\lambda} P_\lambda. \tag{S.18}$$

Proof.

1. Since $\pi^d_{S_n}(\sigma^{-1})$ is in $\mathrm{Comm}(\Delta^n_{\mathrm{U}(d)})$ for all $\sigma \in S_n$, we can apply Lemma S.3 to get

$$\begin{aligned}\mathrm{Tr}(E_{\Delta^n_{\mathrm{U}(d)}}(A)\pi^d_{S_n}(\sigma^{-1})) &= \mathrm{Tr}(E_{\Delta^n_{\mathrm{U}(d)}}(A\pi^d_{S_n}(\sigma^{-1}))) \\ &= \mathrm{Tr}(A\pi^d_{S_n}(\sigma^{-1}))\,,\end{aligned} \tag{S.19}$$

which establishes the first statement.

2. Since the commutant is isomorphic to the group ring, it suffices to proof the statement for all $B = \pi^d_{S_n}(\tau)$ with $\tau \in S_n$. In this case, using the cyclicity of the trace for the first equality, we find

$$\begin{aligned}\Phi(\pi^d_{S_n}(\tau)) &= \sum_{\sigma\in S_n} \mathrm{Tr}(\pi^d_{S_n}(\sigma^{-1})\pi^d_{S_n}(\tau))\pi^d_{S_n}(\sigma) \\ &= \sum_{\sigma\in S_n} \mathrm{Tr}(\pi^d_{S_n}(\tau\sigma^{-1}))\pi^d_{S_n}(\sigma) \\ &= \sum_{\sigma\in S_n} \mathrm{Tr}(\pi^d_{S_n}(\sigma^{-1}))\pi^d_{S_n}(\sigma\tau) \\ &= \pi^d_{S_n}(\tau)\sum_{\sigma\in S_n} \mathrm{Tr}(\pi^d_{S_n}(\sigma^{-1}))\pi^d_{S_n}(\sigma).\end{aligned} \tag{S.20}$$

Here we have used that $\pi^d_{S_n}(\tau\sigma) = \pi^d_{S_n}(\sigma)\pi^d_{S_n}(\tau)$ for all $\tau, \sigma \in S_n$.

3. Using Theorem S.1 (Schur-Weyl duality), we can rewrite $\Phi(\mathrm{Id})$ as

$$\begin{aligned}\Phi(\mathrm{Id}) &= \sum_{\sigma\in S_n} \mathrm{Tr}(\pi^d_{S_n}(\sigma^{-1}))\pi^d_{S_n}(\sigma) \\ &= \sum_{\sigma\in S_n}\sum_{\lambda\vdash n, l(\lambda)\leq d} D_\lambda \,\mathrm{Tr}(\pi_\lambda(\sigma^{-1}))\pi^d_{S_n}(\sigma) \\ &= \sum_{\lambda\vdash n, l(\lambda)\leq d} D_\lambda \sum_{\sigma\in S_n} \chi^\lambda(\sigma)\pi^d_{S_n}(\sigma).\end{aligned} \tag{S.21}$$

The explicit expression (S.8) for the projectors identifies $\Phi(\mathrm{Id})$ as

$$\Phi(\mathrm{Id}) = n! \sum_{\lambda\vdash n, l(\lambda)\leq d} \frac{D_\lambda}{d_\lambda} P_\lambda. \tag{S.22}$$

Since the $\{P_\lambda\}$ are a complete set of orthogonal projectors, the inverse of $\Phi(\mathrm{Id})$ is given by

$$\Phi(\mathrm{Id})^{-1} = \frac{1}{n!}\sum_{\lambda\vdash n, l(\lambda)\leq d} \frac{d_\lambda}{D_\lambda} P_\lambda. \tag{S.23}$$

□

We are now in position to give a concise proof of Theorem S.2:

Proof of Theorem S.2. From Eqns. (S.16) and (S.17) we conclude $\Phi(A) = \Phi(E_{\Delta^n_{\mathrm{U}(d)}}(A)) = E_{\Delta^n_{\mathrm{U}(d)}}(A)\Phi(\mathrm{Id})$ and, thus, $E_{\Delta^n_{\mathrm{U}(d)}}(A) = \Phi(A)\Phi(\mathrm{Id})^{-1}$. Inserting the expression (S.18) for $\Phi(\mathrm{Id})^{-1}$ and the definition (S.15) of Φ yields the expression of the theorem. □

2. *Integration over the Clifford group*

We now turn our attention to the Clifford group and aim at an analogous result to Theorem S.2 for $E_{\Delta^4_{\mathrm{Cl}(d)}}(A)$ with $A \in \mathrm{L}(V)$. As the former result for the unitary group, the result for the Clifford group heavily relies on a characterisation of the commutant of $\Delta^4_{\mathrm{Cl}(d)}$. The required results for the Clifford group were derived in Ref. [10] and apply to multi-qubit dimensions $d = 2^n$. This paper introduces the orthogonal projection

$$Q = \frac{1}{d^2}\sum_{k=1}^{d^2} W_k^{\otimes 4} \tag{S.24}$$

where $W_1, \ldots, W_{d^2} \in \mathrm{L}\left(\mathbb{C}^d\right)$ are the multi-qubit Pauli matrices. In fact, the d^2-dimensional range of Q forms a particular stabiliser code. We denote by $Q^\perp = \mathrm{Id} - Q$ the orthogonal projection onto the complement of this stabiliser code. The orthogonal projection Q commutes with every $\pi^d_{S_4}(\sigma)$, $\sigma \in S_4$. Thus, Q acts trivially on the Specht modules S_λ in the Schur-Weyl decomposition (S.7). Following the notation conventions from Ref. [10], we denote the subspace of the Weyl module W_λ that intersects with the range of Q by W_λ^+ and its dimension as D_λ^+. Analogously, the orthogonal complement of W_λ^+ shall be W_λ^- with dimension D_λ^-. We are now ready to state the main result of this section.

Theorem S.5 (Integration over the Clifford group $\mathrm{Cl}(d)$)**.** *Let $A \in \mathrm{L}(V)$. Then,*

$$\begin{aligned} E_{\Delta^4_{\mathrm{Cl}(d)}}(A) = \frac{1}{4!} \sum_{\lambda \vdash 4, l(\lambda) \leq d} d_\lambda \sum_{\sigma \in S_4} \\ \times \left[\frac{1}{D_\lambda^+} \operatorname{Tr}(A Q \pi^d_{S_4}(\sigma^{-1})) Q \right. \\ \left. + \frac{1}{D_\lambda^-} \operatorname{Tr}(A Q^\perp \pi^d_{S_4}(\sigma^{-1})) Q^\perp \right] \\ \times \pi^d_{S_4}(\sigma) P_\lambda. \end{aligned} \tag{S.25}$$

To set-up the proof we summarise the necessary results of Ref. [10] in the following theorem:

Theorem S.6 (Representation theory of the Clifford group [10])**.** *Whenever $W_\lambda^\pm$ are non-trivial, the action of $\mathrm{Cl}(d) \times S_4$ on $(\mathbb{C}^d)^{\otimes 4}$ is multiplicity free and $(\mathbb{C}^d)^{\otimes 4}$ decomposes into irreducible components*

$$(\mathbb{C}^d)^{\otimes 4} \cong \bigoplus_{\lambda \vdash 4, l(\lambda) \leq d} (W_\lambda^+ \otimes S_\lambda) \oplus (W_\lambda^- \otimes S_\lambda), \tag{S.26}$$

on which $\mathrm{Cl}(d) \times S_4$ acts as $\Delta^\lambda_{\mathrm{Cl}(d)} \otimes \pi^\lambda_{S_4}$.

The dimensions of W_λ^+ are of polynomials in d of degree 4 and the dimensions of W_λ^- are either vanishing or polynomials in d of degree 2.

From Theorem S.6 we learn that an element of the commutant of the diagonal action of the Clifford group $\Delta^4_{\mathrm{Cl}(d)}$ can be written in the form

$$B = Q \bigoplus_{\lambda \vdash 4, l(\lambda) \leq d} (\mathrm{Id}_{D_\lambda} \otimes B_\lambda^+) + Q^\perp \bigoplus_{\lambda \vdash 4, l(\lambda) \leq d} (\mathrm{Id}_{D_\lambda} \otimes B_\lambda^-), \tag{S.27}$$

where $B_\lambda^\pm \in \mathrm{L}(S_\lambda)$ are linear operators acting on the Specht modules S_λ.

To expand elements of $\mathrm{Comm}(\Delta^4_{\mathrm{Cl}(d)})$, we define the map $\tilde{\Phi} : \mathrm{L}(V) \to \mathrm{L}(V)$, $\tilde{\Phi}(A) = \Phi(AQ)Q + \Phi(AQ^\perp)Q^\perp$ with Φ from (S.15). The map $\tilde{\Phi}$ has properties comparable to the map Φ, but is adapted to the diagonal representation of the Clifford group.

Lemma S.7. *For all $A \in \mathrm{L}(V)$ and $B \in \mathrm{Comm}(\Delta^4_{\mathrm{Cl}(d)})$*

$$\tilde{\Phi}(A) = \tilde{\Phi}(E_{\Delta^4_{\mathrm{Cl}(d)}}(A)), \tag{S.28}$$

$$\tilde{\Phi}(B) = B\tilde{\Phi}(\mathrm{Id}), \tag{S.29}$$

$$\tilde{\Phi}(\mathrm{Id})^{-1} = \frac{1}{4!} \sum_{\lambda \vdash 4, l(\lambda) \leq d} d_\lambda P_\lambda \left[\frac{1}{D_\lambda^+} Q + \frac{1}{D_\lambda^-} Q^\perp \right]. \tag{S.30}$$

Proof.

1. Since $Q\pi^d_{S_4}(\sigma^{-1})$ and $Q^\perp \pi^d_{S_4}(\sigma^{-1})$ are in $\mathrm{Comm}(\Delta^4_{\mathrm{Cl}(d)})$ for all $\sigma \in S_4$, we can again apply Lemma S.3 to get $\operatorname{Tr}(E_{\Delta^4_{\mathrm{Cl}(d)}}(A) Q \pi^d_{S_4}(\sigma^{-1})) = \operatorname{Tr}(E_{\Delta^4_{\mathrm{Cl}(d)}}(A Q \pi^d_{S_4}(\sigma^{-1}))) = \operatorname{Tr}(A Q \pi^d_{S_4}(\sigma^{-1}))$ and likewise for $Q^\perp$ instead of Q. Inserting this in the definition of $\tilde{\Phi}$ yields the first statement.

2. From the expansion of elements $B \in \mathrm{Comm}(\Delta^4_{\mathrm{Cl}(d)})$ in (S.27), we conclude that B can be expressed as $B = QB_1 + Q^\perp B_2$, where B_1 and B_2 are in the group ring $\mathbb{C}[S_4]$. Hence, it suffices to show the statement, $\tilde{\Phi}(B) = B\tilde{\Phi}(\mathrm{Id})$, for $B = Q\pi^d_{S_4}(\sigma)$ and $B = Q^\perp \pi^d_{S_4}(\sigma)$. In the first case, we find
$$\begin{aligned} \tilde{\Phi}(Q\pi^d_{S_4}(\sigma)) &= \Phi(Q\pi^d_{S_4}(\sigma))Q \\ &= \Phi(Q\,\mathrm{Id})Q\pi^d_{S_4}(\sigma) \\ &= \tilde{\Phi}(\mathrm{Id})Q\pi^d_{S_4}(\sigma), \end{aligned} \tag{S.31}$$
where property (S.12) from Lemma S.3 has been used in the second step. The proof of $Q^\perp$ is analogous.

3. Using the decomposition (S.26) of Theorem S.6, we can calculate
$$\begin{aligned} \tilde{\Phi}(\mathrm{Id}) &= \sum_{\lambda \vdash 4, l(\lambda) \leq d} \sum_{\sigma \in S_4} \chi_{\pi^d_{S_4}}(\sigma^{-1}) \pi^d_{S_4}(\sigma) \\ &\quad \times \left[D_\lambda^+ Q + D_\lambda^- Q_\lambda^\perp \right] \\ &= 4! \sum_\lambda \frac{1}{d_\lambda} P_\lambda \left[D_\lambda^+ Q + D_\lambda^- Q^\perp \right], \end{aligned} \tag{S.32}$$
where the last line follows again from the expression (S.8) for the projectors. Inverting this expression yields
$$\tilde{\Phi}(\mathrm{Id})^{-1} = \frac{1}{4!} \sum_\lambda d_\lambda P_\lambda \left[\frac{1}{D_\lambda^+} Q + \frac{1}{D_\lambda^-} Q^\perp \right]. \tag{S.33}$$

□

With these statements for the Clifford group at hand, we can proceed to prove Theorem S.5.

Proof of Theorem S.5. Eq. (S.28) in Lemma S.7 and S.29 in Lemma S.7 can be combined to conclude $\tilde{\Phi}(A) = \tilde{\Phi}(E_{\Delta^4_{\mathrm{Cl}(d)}}(A)) = E_{\Delta^4_{\mathrm{U}(d)}}(A)\tilde{\Phi}(\mathrm{Id})$ and, thus, $E_{\Delta^4_{\mathrm{Cl}(d)}}(A) = \tilde{\Phi}(A)\tilde{\Phi}(\mathrm{Id})^{-1}$. The expression for $\tilde{\Phi}(\mathrm{Id})^{-1}$ was derived in Lemma S.7, Eq. (S.30). Together with the definition of $\tilde{\Phi}$ the expression of the theorem follows after some simplification. □

B. The second moment

The main result of this section is the following expression for the second moment of $S_{\mathcal{T}}$ defined in Eq. (S.3). We shall use this statement multiple times in the proofs of our main results.

Lemma S.8 (The 2-nd moment for $\mathrm{U}(d)$)**.** *Let $\mathcal{T} : H_d \to H_d$ be a map. Then*

$$\begin{aligned}&\mathbb{E}_{U\sim\mathrm{Haar}(\mathrm{U}(d))}[S_{\mathcal{T}}^2]\\&=\frac{1}{d^2-1}\Big\{d^2\left\|\mathcal{T}\right\|^2+\operatorname{Tr}(\mathcal{T}(\mathrm{Id}))^2\\&\qquad-\frac{1}{d}\left(\|\mathcal{T}(\mathrm{Id})\|_2^2+\left\|\mathcal{T}^\dagger(\mathrm{Id})\right\|_2^2\right)\Big\},\end{aligned}\tag{S.34}$$

for $S_{\mathcal{T}}$ defined in Eq. (S.3).

For trace-annihilating and Id-annihilating maps, one arrives at a much simpler expression:

Corollary S.9 (Expression for trace-annihilating and Id-annihilating maps)**.** *Let $\mathcal{T} \in \mathrm{V}_{\mathrm{u,tp},0}$ be a map that is trace-annihilating and* Id*-annihilating. Then the second moment of $S_{\mathcal{T}}$ is*

$$\mathbb{E}_{U\sim\mathrm{Haar}(\mathrm{U}(d))}[S_{\mathcal{T}}^2]=\frac{d^2}{d^2-1}\left\|\mathcal{T}\right\|^2.\tag{S.35}$$

Proof. This follows directly from Lemma S.8 and the observation that $\mathcal{T}$ being trace-annihilating translates to $\operatorname{Tr}(\mathcal{T}(\mathrm{Id}))) = 0$ and $\left\|\mathcal{T}^\dagger(\mathrm{Id})\right\|_2 = 0$ and $\mathcal{T}$ being Id-annihilating further requires $\|\mathcal{T}(\mathrm{Id})\|_2 = 0$. $\square$

Before proving Lemma S.8, we derive a general expression for the k-th moment of $S_{\mathcal{T}}$. To this end, recall that by Choi's theorem an endomorphism $\mathcal{T}$ of H_d (i.e. a hermiticity preserving map) can be decomposed as

$$\mathcal{T}(X)=\sum_{i=1}^{r}\lambda_i T_i X T_i^\dagger,\tag{S.36}$$

where $\lambda_i \in \mathbb{R}$ and $T_1,\dots,T_r$ are linear operators with unit Frobenius norm. In this decomposition, the random variable $S_{\mathcal{T}}$ from Eq. (S.3), with $\mathcal{U}(X) = UXU^\dagger$ takes the form

$$S_{\mathcal{T}}=d^2(\mathcal{T},\mathcal{U})=\sum_{i=1}^{r}\lambda_i|\operatorname{Tr}(U^\dagger T_i)|^2\tag{S.37}$$

and its k-th moment can be expressed as follows:

Lemma S.10 (k-th moment of $S_{\mathcal{T}}$)**.** *For $k\in\mathbb{N}$ and T_i defined by Eq.* (S.36) *we have*

$$\begin{aligned}&\mathbb{E}_{U\sim\mathrm{Haar}(\mathrm{U}(d))}[S_{\mathcal{T}}^k]\\&=\sum_{i_1,\dots,i_k=1}^{r}\lambda_{i_1}\cdots\lambda_{i_k}\frac{1}{k!}\sum_{\tau\in S_k}\sum_{\lambda\vdash k,\ l(\lambda)\le d}\frac{d_\lambda}{D_\lambda}\\&\quad\times\operatorname{Tr}\left[\bigotimes_{j=1}^{k}T_{i_{\tau(j)}}^\dagger P_\lambda\bigotimes_{j=1}^{k}T_{i_j}\right].\end{aligned}\tag{S.38}$$

Proof. We can rewrite the k-th unitary moment of $S_{\mathcal{T}}$ as

$$\begin{aligned}&\mathbb{E}_{U\sim\mathrm{Haar}(\mathrm{U}(d))}[S_{\mathcal{T}}^k]\\&=\mathbb{E}_U\sum_{i_1,\dots,i_k=1}^{r}\lambda_{i_1}\cdots\lambda_{i_k}|\operatorname{Tr}(U^\dagger T_{i_1})|^2\cdots|\operatorname{Tr}(U^\dagger T_{i_k})|^2\\&=\mathbb{E}_U\sum_{i_1,\dots,i_k=1}^{r}\lambda_{i_1}\cdots\lambda_{i_k}\\&\quad\times\operatorname{Tr}\left[\bigotimes_{j=1}^{k}T_{i_j}^\dagger\,U^{\otimes k}\right]\operatorname{Tr}\left[U^{\dagger\otimes k}\bigotimes_{j=1}^{k}T_{i_j}\right]\\&=\sum_{i_1,\dots,i_k=1}^{r}\lambda_{i_1}\cdots\lambda_{i_k}\\&\quad\times\sum_{m,n=1}^{d^k}\langle m|\bigotimes_{j=1}^{k}T_{i_j}^\dagger E_{\Delta^k_{\mathrm{U}(d)}}(|m\rangle\langle n|)\bigotimes_{j=1}^{k}T_{i_j}\,|n\rangle\end{aligned}\tag{S.39}$$

where in the last line we evaluated the trace in an orthonormal basis $\{|m\rangle \mid m\in\{1,\dots,d^k\}\}$ for $(\mathbb{C}^d)^{\otimes k}$. Using the expression for $E_{\Delta^k_{\mathrm{U}(d)}}$ of Theorem S.2 we get

$$\begin{aligned}&\mathbb{E}_{U\sim\mathrm{Haar}(\mathrm{U}(d))}[S_{\mathcal{T}}^k]\\&=\sum_{i_1,\dots,i_k=1}^{r}\lambda_{i_1}\cdots\lambda_{i_k}\frac{1}{k!}\sum_{\tau\in S_k}\sum_{\lambda\vdash k,\ l(\lambda)\le d}\frac{d_\lambda}{D_\lambda}\\&\quad\times\operatorname{Tr}\left[\pi^d_{S_k}(\tau)\bigotimes_{j=1}^{k}T_{i_j}^\dagger\pi^d_{S_k}(\tau^{-1})P_\lambda\bigotimes_{j=1}^{k}T_{i_j}\right]\\&=\sum_{i_1,\dots,i_k=1}^{r}\lambda_{i_1}\cdots\lambda_{i_k}\frac{1}{k!}\sum_{\tau\in S_k}\sum_{\lambda\vdash k,\ l(\lambda)\le d}\frac{d_\lambda}{D_\lambda}\\&\quad\times\operatorname{Tr}\left[\bigotimes_{j=1}^{k}T_{i_{\tau(j)}}^\dagger P_\lambda\bigotimes_{j=1}^{k}T_{i_j}\right].\end{aligned}\tag{S.40}$$

$\square$

Proof of Lemma S.8. We evaluate the expression of Lemma S.10 for the case $k = 2$. To this end recall that the irreducible representations of S_2 are the symmetric ($\square\!\square$) and antisymmetric representation ($\begin{smallmatrix}\square\\\square\end{smallmatrix}$). The central projections are given by $P_{\square\!\square} = \frac{1}{2}(1+\mathbb{F})$ and $P_{\begin{smallmatrix}\square\\\square\end{smallmatrix}} = \frac{1}{2}(1-\mathbb{F})$ [7], where $\mathbb{F}$ is the bipartite flip operator $\mathbb{F} : (\mathbb{C}^d)^{\otimes 2} \to (\mathbb{C}^d)^{\otimes 2}$, $|x\rangle\otimes|y\rangle \mapsto |y\rangle\otimes|x\rangle$. The dimensions are $d_{\square\!\square} = d_{\begin{smallmatrix}\square\\\square\end{smallmatrix}} = 1$, $D_{\square\!\square} = \frac{d(d-1)}{2}$ and $D_{\begin{smallmatrix}\square\\\square\end{smallmatrix}} = \frac{d(d+1)}{2}$. For $A, B \in H_d^{\otimes 2}$ we introduce the following short-hand notation

$$\Gamma_{AB} := \sum_{i,j}^{r}\lambda_i\lambda_j\operatorname{Tr}\left[A(T_i^\dagger\otimes T_j^\dagger)B(T_i\otimes T_j)\right].\tag{S.41}$$

Rearranging the terms in the first statement of the Lemma S.10

then yields

$$
\begin{aligned}
&\mathbb{E}_{U\sim\mathrm{Haar}(\mathrm{U}(d))}[S_{\mathcal{T}}^2] && \text{(S.42)}\\
&= \frac{1}{4}\left\{ \left[\frac{1}{D_{\square\square}} + \frac{1}{D_{\boxminus}}\right] [\Gamma_{\mathrm{Id}\,\mathrm{Id}} + \Gamma_{\mathbb{F}\mathbb{F}}] \right. && \text{(S.43)}\\
&\qquad \left. + \left[\frac{1}{D_{\square\square}} - \frac{1}{D_{\boxminus}}\right] [\Gamma_{\mathbb{F}\,\mathrm{Id}} + \Gamma_{\mathrm{Id}\,\mathbb{F}}] \right\} && \text{(S.44)}\\
&= \frac{1}{d^2-1}\left\{\Gamma_{\mathrm{Id}\,\mathrm{Id}} + \Gamma_{\mathbb{F}\mathbb{F}} - \frac{1}{d}\left(\Gamma_{\mathrm{Id}\,\mathbb{F}} + \Gamma_{\mathbb{F}\,\mathrm{Id}}\right)\right\}. && \text{(S.45)}
\end{aligned}
$$

The four Γ-terms can be evaluated explicitly. For the first term, we obtain

$$
\begin{aligned}
\Gamma_{\mathrm{Id}\,\mathrm{Id}} &= \sum_{i,j=1}^{r} \lambda_i\lambda_j \left\|T_i\right\|_2^2 \left\|T_j\right\|_2^2 \\
&= \left(\sum_i \lambda_i \operatorname{Tr}(T_i \,\mathrm{Id}\, T_i^\dagger)\right)^2 \\
&= \operatorname{Tr}(\mathcal{T}(\mathrm{Id}))^2 .
\end{aligned} \tag{S.46}
$$

The second terms reads

$$
\begin{aligned}
\Gamma_{\mathbb{F}\mathbb{F}} &= \sum_{i,j=1}^{r} \lambda_i\lambda_j |\operatorname{Tr}(T_i^\dagger T_j)|^2 \\
&= d^2 \left\|\mathcal{T}\right\|^2
\end{aligned} \tag{S.47}
$$

and the third term can be written as

$$
\begin{aligned}
\Gamma_{\mathbb{F}\,\mathrm{Id}} &= \sum_{i,j=1}^{r} \lambda_i\lambda_j \operatorname{Tr}\left(T_i^\dagger T_i T_j^\dagger T_j\right) \\
&= \left\|\mathcal{T}^\dagger(\mathrm{Id})\right\|_2^2 .
\end{aligned} \tag{S.48}
$$

Moreover, a computation that closely resembles this reformulation yields $\Gamma_{\mathrm{Id}\,\mathbb{F}} = \left\|\mathcal{T}(\mathrm{Id})\right\|_2^2$ and the claim follows. □

C. A fourth moment bound

The main result of this section is an upper bound for the fourth moment of $S_{\mathcal{T}}$ when $\mathcal{U}$ is a Clifford operation drawn uniformly at random. To gain some intuition, let us first derive an upper bound on the fourth moment taken with respect to the full unitary group. Note that a similar bound has already been derived in Ref. [11].

Lemma S.11 (4-th moment bound for $\mathrm{U}(d)$)**.** *Let $\mathcal{T} : H_d \to H_d$ be a map. Then for $S_{\mathcal{T}}$ defined in Eq. (S.3)*

$$
\mathbb{E}_{U\sim\mathrm{Haar}(\mathrm{U}(d))}[S_{\mathcal{T}}^4] \le C \left\|J(\mathcal{T})\right\|_1^4 \tag{S.49}
$$

with some constant $C > \frac{1}{3}$ independent of the dimension d.

Proof. Applying Cauchy-Schwarz to an individual summand on the right hand side of Lemma S.10 yields for all k

$$
\begin{aligned}
\left|\operatorname{Tr}\left[\bigotimes_{j=1}^{k} T_{i_{\tau(j)}}^\dagger P_\lambda \bigotimes_{j=1}^{k} T_{i_j}\right]\right| &\le \left\|P_\lambda \bigotimes_{j=1}^{k} T_{i_{\tau(j)}}\right\|_2 \left\|P_\lambda \bigotimes_{j=1}^{k} T_{i_j}\right\|_2 \\
&\le \left\|\bigotimes_{j=1}^{k} T_{i_{\tau(j)}}\right\|_2 \left\|\bigotimes_{j=1}^{k} T_{i_j}\right\|_2 \\
&= \prod_{j=1}^{k} \left\|T_{i_j}\right\|_2^2 ,
\end{aligned} \tag{S.50}
$$

which is independent of the permutation $\tau \in S_k$. We may therefore conclude

$$
\begin{aligned}
&\mathbb{E}_{U\sim\mathrm{Haar}(\mathrm{U}(d))}[S_{\mathcal{T}}^k] \\
&\le \sum_{i_1,\dots,i_k=1}^{r} \prod_{j=1}^{k} |\lambda_{i_j}| \left\|T_{i_j}\right\|_2^2 \sum_{\lambda\vdash k,\ l(\lambda)\le d} \frac{d_\lambda}{D_\lambda}.
\end{aligned} \tag{S.51}
$$

From Theorem S.6 we observe that for $k = 4$

$$
\sum_{\lambda\vdash 4,\ l(\lambda)\le d} \frac{d_\lambda}{D_\lambda} \le \frac{C}{d^4} \tag{S.52}
$$

for some constant $C > \frac{1}{3}$ independent of d. Thus, Eq. (S.51) implies the desired bound. □

In an analogous way we can derive a sufficient bound on the fourth moment of $S_{\mathcal{T}}$ when the average is performed over the Clifford group. The result will be stated in Lemma S.15. To get the correct dimensional pre-factors in the bound, we have to rely on particular properties of the projection Q of Eq. (S.24) appearing in the representation theory of the fourth order diagonal action of Clifford group in Theorem S.5. The following technical result takes care of this issue.

Lemma S.12 (Properties of the projection Q)**.** *For $\{T_l\}_{l=1}^r \subset \mathrm{L}(\mathbb{C}^d)$ and Q defined in Eq. (S.24)*

$$
\left\|Q \bigotimes_{j=1}^{4} T_{i_j} Q\right\|_2 \le \frac{1}{d} \prod_{j=1}^{4} \left\|T_{i_j}\right\|_2 . \tag{S.53}
$$

This bound is tight. In fact, one can show that it is saturated if all T_i's are chosen to be the same stabiliser state. The proof of Lemma S.12 requires two other properties of multi-qubit Pauli matrices $W_1, \dots, W_{d^2}$. The first property is summarised by the following lemma.

Lemma S.13 (Magnitude of multi-qubit Pauli matrices)**.** *For $A, B \in \mathrm{L}(\mathbb{C}^d)$,*

$$
\operatorname{Tr}(W_j A W_k B) \le \left\|A\right\|_2 \left\|B\right\|_2 \tag{S.54}
$$

for all $j, k \in \{1, \dots, d^2\}$.

Proof. This statement follows directly from Cauchy-Schwarz and the unitary invariance of the Frobenius norm:

$$\begin{aligned}\operatorname{Tr}\left(W_j A W_k B\right) &= \left(B^\dagger, W_j A W_k\right) \\ &\leq \|B^\dagger\|_2 \|W_j A W_k\|_2 \\ &= \|B\|_2 \|A\|_2.\end{aligned} \tag{S.55}$$

□

The second property is that the two multi-qubit flip operator $\mathbb{F}$ can be expanded in terms of tensor products of Pauli matrices.

Lemma S.14 (Multi-qubit flip operator in terms of Pauli matrices)**.**

$$\mathbb{F} = \frac{1}{d}\sum_{i=1}^{d^2} W_i^{\otimes 2}. \tag{S.56}$$

Proof. The re-normalised Pauli matrices form an orthonormal basis of H_d:

$$X = \frac{1}{d}\sum_{k=1}^{d} W_k \operatorname{Tr}\left(W_k X\right) \quad \forall X \in H(\mathbb{C}^n). \tag{S.57}$$

We can extend this to a basis of $H_d^{\otimes 2}$ by considering all possible tensor products of Pauli matrices. Expanding the flip operator in this basis yields

$$\begin{aligned}\mathbb{F} &= \frac{1}{d^2}\sum_{k,l=1}^{d^2} W_k \otimes W_l \operatorname{Tr}\left(\mathbb{F} W_k \otimes W_l\right) \\ &= \frac{1}{d^2}\sum_{k,l=1}^{d^2} W_k \otimes W_l d\delta_{k,l} = \frac{1}{d}\sum_{k=1}^{d^2} W_k^{\otimes 2}\end{aligned} \tag{S.58}$$

as claimed. □

We are now equipped to prove Lemma S.12.

Proof of Lemma S.12. We start by inserting the definition of Q, (S.24). Fixing w.l.o.g. an order of the indices, we obtain

$$\operatorname{Tr}\left[Q \bigotimes_{j=1}^{4} T_j Q \bigotimes_{j=1}^{4} T_j^\dagger\right] \tag{S.59}$$

$$= \frac{1}{d^4}\sum_{k,l=1}^{d^2}\prod_{j=1}^{4}\operatorname{Tr}\left[W_k T_j W_l T_j^\dagger\right] \tag{S.60}$$

$$= \frac{1}{d^4}\sum_{k,l=1}^{d^2} c_{k,l}(T_1)c_{k,l}(T_2)c_{k,l}(T_3)c_{k,l}(T_4), \tag{S.61}$$

where we defined $c_{k,l}(T_j) := \operatorname{Tr}(W_k T_j W_l T_j^\dagger) \in \mathbb{C}$. These numbers obey

$$\begin{aligned}\overline{c_{k,l}(T_j)} &= \overline{\operatorname{Tr}\left(W_k T_j W_l T_j^\dagger\right)} = \operatorname{Tr}\left(\left(W_k T_j W_l T_j^\dagger\right)^\dagger\right) \\ &= \operatorname{Tr}\left(T_j W_l^\dagger T_j^\dagger W_k\right) = c_{k,l}(T_j^\dagger).\end{aligned} \tag{S.62}$$

In addition, Lemma S.13 implies

$$|c_{k,l}(T_j)|^2 = \left|\operatorname{Tr}\left(W_k T_j W_l T_j^\dagger\right)\right|^2 \leq \|T_j\|_2^4. \tag{S.63}$$

Equation (S.61) can be viewed as a complex-valued inner product between two d^2-dimensional vectors indexed by k and l. This expression can be upper bounded by the Cauchy-Schwarz inequality:

$$\frac{1}{d^4}\sum_{k,l=1}^{d^2} c_{k,l}(T_1)c_{k,l}(T_2)c_{k,l}(T_3)c_{k,l}(T_4) \tag{S.64}$$

$$= \frac{1}{d^4}\sum_{k,l=1}^{d^2} \overline{c_{k,l}(T_1^\dagger)c_{k,l}(T_2^\dagger)}c_{k,l}(T_3)c_{k,l}(T_4) \tag{S.65}$$

$$\begin{aligned}&\leq \frac{1}{d^2}\sqrt{\frac{1}{d^2}\sum_{k,l}\left|c_{k,l}(T_1^\dagger)c_{k,l}(T_2^\dagger)\right|^2} \\ &\quad\times\sqrt{\frac{1}{d^2}\sum_{k,l}\left|c_{k,l}(T_3)c_{k,l}(T_4)\right|^2}.\end{aligned} \tag{S.66}$$

The first square-root can be bounded in the following way

$$\begin{aligned}&\sqrt{\frac{1}{d^2}\sum_{k,l}\left|c_{k,l}(T_3)c_{k,l}(T_4)\right|^2} \\ &\leq \sqrt{\|T_1^\dagger\|_2^4 \frac{1}{d^2}\sum_{k,l} c_{k,l}(T_2^\dagger)} \\ &= \|T_1\|_2^2\sqrt{\frac{1}{d^2}\sum_{k,l}\operatorname{Tr}\left(W_k T_2^\dagger W_l T_2\right)^2} \\ &= \|T_1\|_2^2\sqrt{\operatorname{Tr}\left(\frac{1}{d}\sum_k W_k^{\otimes 2}(T_2^\dagger)^{\otimes 2}\frac{1}{d}\sum_l W_l^{\otimes 2}T_2^{\otimes 2}\right)} \\ &= \|T_1\|_2^2\sqrt{\operatorname{Tr}\left(\mathbb{F}\,(T_2^\dagger)^{\otimes 2}\,\mathbb{F}\,T_2^{\otimes 2}\right)} \\ &= \|T_1\|_2^2\sqrt{\operatorname{Tr}\left(T_2^\dagger T_2\right)^2} = \|T_1\|_2^2\|T_2\|_2^2.\end{aligned} \tag{S.67}$$

Here, we have applied the magnitude bound (S.63) for $c_{k,l}(T_1^\dagger)$ in the second line and applied Lemma S.14.

The second square root can be bounded in a complete analogous fashion, i.e.

$$\sqrt{\frac{1}{d^2}\sum_{k,l}\left|c_{k,l}(T_3)c_{k,l}(T_4)\right|^2} \leq \|T_3\|_2^2\|T_4\|_2^2. \tag{S.68}$$

Inserting both bounds into Eq. (S.66) yields the desired claim. □

Having established Lemma S.12, we will now state the bound on the fourth moment of $S_{\mathcal{T}}$ when the average is performed over the Clifford group.

Lemma S.15 (4-th moment bound for $\mathrm{Cl}(d)$). *Let $\mathcal{T} : H_d \to H_d$ be a map. For $S_{\mathcal{T}}$ defined in Eq. (S.3), it holds*

$$\mathbb{E}_{U\sim \mathrm{Haar}(\mathrm{Cl}(d))}[S_{\mathcal{T}}^4] \le C\,\|J(\mathcal{T})\|_1^4, \tag{S.69}$$

where $\|\cdot\|_1$ denotes the trace (or nuclear) norm and the constant $C>0$ is independent of d.

Proof. As for the unitary group, we can rewrite the k-th moment of $S_{\mathcal{T}}$ for the Clifford group as

$$\begin{aligned}&\mathbb{E}_{U\sim \mathrm{Haar}(\mathrm{Cl}(d))}[S_{\mathcal{T}}^k]\\ &= \sum_{i_1,\dots,i_k=1}^{r} \lambda_{i_1}\cdots\lambda_{i_k} \sum_{m,n=1}^{d^k}\\ &\times \langle m|\bigotimes_{j=1}^{k} T_{i_j}^\dagger E_{\Delta^k_{\mathrm{Cl}(d)}}(|m\rangle\langle n|)\bigotimes_{j=1}^{k} T_{i_j}\,|n\rangle\end{aligned} \tag{S.70}$$

using a basis $\{|m\rangle \mid m\in\{1,\dots,d^k\}\}$ for $(\mathbb{C}^d)^{\otimes k}$. The expression for $E_{\Delta^4_{\mathrm{Cl}(d)}}$ with $k=4$ was derived in Theorem S.5. It implies that

$$\begin{aligned}&\mathbb{E}_{U\sim \mathrm{Haar}(\mathrm{Cl}(d))}[S_{\mathcal{T}}^4]\\ &= \sum_{i_1,\dots,i_k=1}^{r} \lambda_{i_1}\cdots\lambda_{i_k}\frac{1}{4!}\sum_{\tau\in S_k}\sum_{\lambda\vdash k,\ l(\lambda)\le d} d_\lambda\\ &\times\Bigg\{\frac{1}{D_\lambda^+}\mathrm{Tr}\left[Q\bigotimes_{j=1}^{4}T_{i_{\tau(j)}}^\dagger Q P_\lambda \bigotimes_{j=1}^{4}T_{i_j}\right]\\ &+\frac{1}{D_\lambda^-}\mathrm{Tr}\left[Q^\perp\bigotimes_{j=1}^{4}T_{i_{\tau(j)}}^\dagger Q^\perp P_\lambda \bigotimes_{j=1}^{4}T_{i_j}\right]\Bigg\}.\end{aligned} \tag{S.71}$$

We may bound the first trace term by

$$\begin{aligned}&\left|\mathrm{Tr}\left[Q\bigotimes_{j=1}^{4}T_{i_{\tau(j)}}^\dagger Q P_\lambda \bigotimes_{j=1}^{4}T_{i_j}\right]\right|\\ &\le \left\|P_\lambda Q\bigotimes_{j=1}^{4}T_{i_{\tau(j)}}Q\right\|_2\left\|P_\lambda Q\bigotimes_{j=1}^{4}T_{i_j}Q\right\|_2\\ &\le \left\|Q\bigotimes_{j=1}^{4}T_{i_{\tau(j)}}Q\right\|_2\left\|Q\bigotimes_{j=1}^{4}T_{i_j}Q\right\|_2\\ &\le \frac{1}{d^2}\prod_{j=1}^{4}\|T_{i_j}\|_2^2,\end{aligned} \tag{S.72}$$

where we have used Cauchy-Schwarz and applied Lemma S.12 in the last line. For the second trace term a looser bound suffices:

$$\left\|Q^\perp\bigotimes_{j=1}^{k}T_{i_{\tau(j)}}Q^\perp\right\|_2 \le \prod_{j=1}^{k}\|T_{i_j}\|_2 \tag{S.73}$$

for all $\tau\in S_4$. This follows directly from Cauchy-Schwarz. Altogether we conclude that

$$\begin{aligned}&\mathbb{E}_{U\sim \mathrm{Haar}(\mathrm{Cl}(d))}[S_{\mathcal{T}}^4]\\ &\le \sum_{i_1,\dots,i_4=1}^{r}\prod_{j=1}^{4}|\lambda_{i_j}|\,\|T_{i_j}\|_2^2\sum_{\lambda\vdash k,\ l(\lambda)\le d} d_\lambda\left[\frac{1}{d^2 D_\lambda^+}+\frac{1}{D_\lambda^-}\right]\\ &\le C\,\|J(\mathcal{T})\|_1^4\end{aligned} \tag{S.74}$$

with some constant $C>0$ independent of d. The last step follows from the dimensions given in Theorem S.6. ☐

D. Proof of Theorem 2 (recovery guarantee)

We consider the following measurements: For a map $\mathcal{X}\in \mathrm{L}(H_d)$ the measurement outcomes $f\in\mathbb{R}^m$ are given by

$$\begin{aligned}f_i &= F_{\mathrm{avg}}(\mathcal{C}_i,\mathcal{X})+\epsilon_i\\ &= \frac{1}{d+1}\left[d\,(\mathcal{C}_i,\mathcal{X})+\frac{1}{d}\mathrm{Tr}(\mathcal{X}^\dagger(\mathrm{Id}))\right]+\epsilon_i,\end{aligned} \tag{S.75}$$

where $\mathcal{C}_i$ are random Clifford channels and $\epsilon\in\mathbb{R}^m$ accounts for additional additive noise.

To make use of the proof techniques developed for low rank matrix reconstruction [12, 13], we will in the following work in the Choi representation of channels. This has the advantage, that the Kraus rank directly translates to the familiar matrix rank. We define the Choi matrix of a map $\mathcal{X}\in\mathrm{L}(H_d)$ as

$$J(\mathcal{X}) = (\mathcal{X}\otimes\mathrm{Id})(|\psi\rangle\langle\psi|), \tag{S.76}$$

where $|\psi\rangle = d^{-1/2}\sum_{k=1}^{d}|k\rangle\otimes|k\rangle\in\mathbb{C}^d\otimes\mathbb{C}^d$ is the maximally entangled state vector. The Choi matrix of a map is positive semi-definite if and only if the map is completely positive. We denote the cone of positive semi-definite matrices by Pos_{d^2}. A channel $\mathcal{X}$ is trace-preserving and unital if and only if both partial traces of the Choi matrix yield the maximally mixed state, i.e. $\mathrm{Tr}_1(J(\mathcal{X}))=\mathrm{Tr}_2(J(\mathcal{X}))=\mathrm{Id}/d$. We will denote the set of Choi matrices that correspond to channels in $\mathrm{L}_{\mathrm{u,tp}}$ by $J(\mathrm{L}_{\mathrm{u,tp}})$. Furthermore, we define $J(\mathrm{V}_{\mathrm{u,tp,0}})$ as the set of Choi matrices corresponding to trace- and identity-annihilating channels, i.e., both partial traces of operators in $J(\mathrm{V}_{\mathrm{u,tp,0}})$ vanish. Moreover, recall that the inner product on $\mathrm{L}_{\mathrm{u,tp}}$ we introduced in (S.1) coincides with the Hilbert-Schmidt inner product of the corresponding Choi matrices (S.2). Adhering to this correspondence, we slightly abuse notation and use $(\mathcal{X},\mathcal{Y})$ and $(J(\mathcal{X}),J(\mathcal{Y}))$ interchangeably.

To formalise the robustness of our reconstruction we need to introduce the following notation. For a Hermitian matrix $Z\in H_d$ let λ be the largest eigenvalue with an eigenvector v. We write $Z|_1=\lambda\,|v\rangle\langle v|$ for the best unit rank approximation to Z and $Z|_c := Z - Z|_1$ denotes the corresponding "tail".

In terms of the Choi matrix of $\mathcal{X}$ the measurement outcomes $f\in\mathbb{R}^m$ read

$$f_i = \frac{1}{d+1}\left[d\,(J(\mathcal{C}_i),J(\mathcal{X}))+\mathrm{Tr}(J(\mathcal{X}))\right]+\epsilon_i, \tag{S.77}$$

The underlying linear measurement map $\mathcal{A} : H_{d^2} \to \mathbb{R}^m$ is given by

$$\mathcal{A}_i(X) = \frac{1}{d+1}\left[d(J(\mathcal{C}_i), X) + \mathrm{Tr}(X)\right]. \tag{S.78}$$

Since unital and trace preserving maps $\mathcal{X}$ have trace normalised Choi matrices the second trace-term of the measurement map is just a constant shift. We also define the set of measurement matrices $\{A_i\}_{i=1}^{b}$ that encode the measurement map as $\mathcal{A}_i(X) = (A_i, X)$: $A_i = \frac{d}{d+1}\left[J(\mathcal{C}_i) + \mathrm{Id}/d\right]$, where each $\mathcal{C}_i$ is a gate that is chosen uniformly at random from the multi-qubit Clifford group.

In the Choi representation, we want to consider the optimisation problem

$$\begin{aligned} \underset{Z}{\text{minimise}} \quad & \|\mathcal{A}(Z) - f\|_{\ell_q} \\ \text{subject to} \quad & Z \in J(\mathrm{L}_{\text{u,tp}}) \cap \mathrm{Pos}_{d^2}, \end{aligned} \tag{S.79}$$

where we allow the minimisation of an arbitrary ℓ_q-norm. The optimisation problem (3) is equivalent to (S.79) for $q = 2$.

We are interested in using the optimisation procedure (S.79) for the recovery of unitary quantum channels. In this section, we will derive the following recovery guarantee:

Theorem S.16 (Recovery guarantee)**.** *Let $\mathcal{A} : H_{d^2} \to \mathbb{R}^m$ be the measurement map* (S.78) *with*

$$m \geq c d^2 \log(d). \tag{S.80}$$

Then, for all $X \in J(L_{\text{u,tp}}) \cap \mathrm{Pos}_{d^2}$ given noisy observations $f = \mathcal{A}(X) + \epsilon \in \mathbb{R}^m$, the minimiser $Z^\sharp$ of the optimisation problem (S.79) *fulfils for $p \in \{1, 2\}$*

$$\left\|Z^\sharp - X\right\|_p \leq \tilde{C}_1 \left\|X|_c\right\|_1 + 2\tilde{C}_2 d^2 m^{-1/q} \|\epsilon\|_{\ell_q} \tag{S.81}$$

with probability at least $1 - \mathrm{e}^{-c_f m}$ over the random measurements. The constants $\tilde{C}_1, \tilde{C}_2, c, c_f > 0$ only depend on each other.

The recovery guarantee of Theorem 2 is the special case of Theorem S.16 for $q = 2$ and $p = 2$ restricted to measurements of a unitary quantum channel. In contrast, the more general formulation of Theorem S.16 allows for a violation of the unit rank assumption. The first term (S.81) is meant to absorb violations of this assumption into the error bound. We note in passing that the choice of $p = 1$ actually yields a tighter bound compared to $p = 2$.

More generally, one can ask for a recovery guarantee if the measured map X can not be guaranteed to be unital or trace preserving. From Eq. S.164 one observes that as long as the map X is trace normalised the measured AGFs are identical to the average fidelities of the projection $X_{\text{u,tp}}$ of X onto the affine space of unital and trace-preserving maps. But since $X_{\text{u,tp}}$ is not necessarily positive, it is not straight-forward to apply Theorem S.16 to $X_{\text{u,tp}}$. We expect the reconstruction algorithm to recover the trace-preserving and unital part of an arbitrary map. The reconstruction error (S.81) is expected to additionally feature a term proportional to the distance of X to the intersection of $L_{\text{u,tp}}$ with the cone Pos_{d^2} of positive semi-definite matrices.

Another way to proceed is to use a trace-norm minimisation subject to unitality, trace-preservation and the data constraints $\|\mathcal{A}(Z) - f\|_{\ell_q} < \eta$. The derivation of Theorem S.16 readily yields a recovery guarantee for the trace-norm minimisation that is essentially identical to Theorem S.16. See Ref. [13] for details on the argument. The main difference is that such a recovery guarantee does not need to assume complete positivity of the map that is to be reconstructed. Correspondingly, the result of the trace-norm minimisation is not guaranteed to be positive semi-definite. This implies that the robustness of this algorithm against violations of the unitality and trace-preservation is different compared to (S.79). For example, the AGFs of a not necessarily unital or trace-preserving map $\mathcal{X}$ to unitary gates coincide with the AGFs of its unital and trace-preserving part $\mathcal{X}_{\text{u,tp}}$ as long as X is still normalised in trace-norm. This is a consequence of Eq. S.164. Thus, a trace-norm minimisation will reconstruct $X_{\text{u,tp}}$ up to an error given by $\|J(\mathcal{X}_{\text{u,tp}})|_c\|_1$ and noise. We leave a more extensive study of the robustness of the discussed reconstruction algorithms against violations of this particular model assumption to future work.

The proof of the recovery guarantee relies on establishing the so-called *null space property (NSP)* for the measurement map $\mathcal{A}$. We refer to Ref. [14] for a history of the term. The NSP ensures injectivity, i.e. informational completeness, of the measurement map $\mathcal{A}$ restricted to the matrices that should be recovered. Informally, for our purposes, a measurement map $\mathcal{A} : H_{d^2} \to \mathbb{R}^m$ obeys the NSP if no unit rank matrix in $J(\mathrm{V}_{\text{u,tp,0}})$ is in the kernel (nullspace) of $\mathcal{A}$.

Definition S.17 (Robust NSP, Definition 3.1 in Ref. [13])**.** *$\mathcal{A} : H_{d^2} \to \mathbb{R}^m$ satisfies the* null space property (NSP) *with respect to ℓ_q with constant $\tau > 0$ if for all $X \in J(\mathrm{V}_{\text{u,tp,0}})$*

$$\|X|_1\|_2 \leq \frac{1}{2}\left\|X|_c\right\|_1 + \tau \|\mathcal{A}(X)\|_{\ell_q}. \tag{S.82}$$

The factor $1/2$ in front of the first term of (S.82) is only one possible choice. In fact, one can instead introduce a constant with value in $(0, 1)$. The constants appearing in Theorem S.16 then depend on the specific value of the pre-factor. In particular, the different choices of the pre-factor in the definition of the NSP result in different trade-offs between the constant c that appears in the sampling complexity and the constant $\tilde{C}_1$ that decorates the model-mismatch term in the reconstruction error. For the simplicity, we leave these dependencies implicit.

The main consequence of the NSP that we require is captured by the following reformulation of Theorem 12 of [13].

Theorem S.18. *Fix $p \in \{1, 2\}$ and let $\mathcal{A} : H_{d^2} \to \mathbb{R}^m$ satisfy the NSP with constant $\tau > 0$. Then, for all $Y, Z \in J(\mathrm{L}_{\text{u,tp}})$*

$$\begin{aligned} \|Z - Y\|_p \leq \frac{9}{2}\left[\|Z\|_1 - \|Y\|_1 + 2\left\|Y|_c\right\|_1\right] \\ + 7\tau \left\|\mathcal{A}(Z - Y)\right\|_{\ell_q}. \end{aligned} \tag{S.83}$$

In fact, the measurement $\mathcal{A}$ of (S.77) obeys the NSP. More precisely:

Lemma S.19. *Let $\mathcal{A} : H_{d^2} \to \mathbb{R}^m$ be the measurement map defined in (S.78) with $m \geq cd^2 \log(d)$. Then $\mathcal{A}$ obeys the NSP property with constant $\tau = C^{-1} d(d+1) m^{-1/q}$ with probability of at least $1 - \mathrm{e}^{-c_f m}$. The constants $C, c, c_f > 0$ only depend on each other.*

The proof of Lemma S.19 is developed in the subsequent section.

Proof of Theorem S.16. With the requirements of Lemma S.19 we can apply Theorem S.18 and set $Z = Z^\sharp$, the reconstructed result of the algorithm, as well as $Y = X$. The theorem's statement then reads

$$\begin{aligned}\left\|Z^\sharp - X\right\|_p &\leq 9 \left\|X|_c\right\|_1 \\ &\quad + 7\tau \left\|\mathcal{A}(Z^\sharp - X)\right\|_{\ell_q},\end{aligned} \tag{S.84}$$

because $\|X\|_1 = \|Z\|_1 = 1$ is true for arbitrary Choi matrices of (trace-preserving) quantum channels. The second term is dominated by

$$\begin{aligned}\left\|\mathcal{A}(Z^\sharp - X)\right\|_{\ell_q} &\leq \left[\left\|\mathcal{A}(X - Z^\sharp) + \epsilon\right\|_{\ell_q} + \|\epsilon\|_{\ell_q}\right] \\ &\leq 2 \|\epsilon\|_{\ell_q},\end{aligned} \tag{S.85}$$

where the last step follows from $Z^\sharp$ being the minimiser of (S.79). Thus, we can replace it by any point in the feasible set including X on the right hand side of the first line. Inserting (S.85) and the NSP constants of Lemma S.19 into (S.84) the assertion of the theorem follows. ☐

In the remainder of this section, we will establish the NSP for our measurement matrix $\mathcal{A}$ as summarised in Lemma S.19.

Establishing the null space property

To prove Lemma S.19 at the end of this section we start with deriving a criterion for the NSP property following the approach taken in Refs. [13, 15].

Lemma S.20. *A map $\mathcal{A} : H_{d^2} \to \mathbb{R}^m$ obeys the null space property with respect to ℓ_q-norm with constant $\tau > 0$ if*

$$\inf_{X \in \mathbf{\Omega}} \|\mathcal{A}(X)\|_{\ell_1} \geq \frac{m^{1-1/q}}{\tau} \tag{S.86}$$

with

$$\mathbf{\Omega} := \{Z \in J(\mathrm{V}_{\mathrm{u,tp,0}}) \mid \|Z|_1\|_2 \geq \frac{1}{2} \|Z|_c\|_1 , \|Z\|_2 = 1\}.$$

Proof. For matrices X with the property $\|X|_1\|_2 \leq \frac{1}{2} \|X|_c\|_1$ the NSP condition (S.82) is satisfied independently of the map $\mathcal{A}$. Hence, to establish the NSP for a specific map $\mathcal{A}$ it suffice to show that the condition (S.82) holds for all $X \in \mathbf{\Omega} = \{Z \in J(\mathrm{V}_{\mathrm{u,tp,0}}) \mid \|Z|_1\|_2 \geq \frac{1}{2} \|Z|_c\|_1 , \|Z\|_2 = 1\}$. The additional assumption of $\|Z\|_2 = 1$ is no restriction since both sides of (S.82) are absolutely homogeneous functions of the same degree. By definition, for all $X \in \mathbf{\Omega}$ we have $\|X|_1\|_2 \leq \|X\|_2 \leq 1$. Therefore, for $X \in \mathbf{\Omega}$

$$\|\mathcal{A}(X)\|_{\ell_q} \geq \frac{1}{\tau} \tag{S.87}$$

implies the NSP condition (S.82). Using the norm inequality $\|x\|_{\ell_q} \geq m^{1/q-1} \|x\|_{\ell_1}$ yields the criterion of the lemma. ☐

Recall that every rank-r matrix X obeys $\|X\|_1^2 / \|X\|_2^2 \leq r$. This motivates thinking of the matrices of $\mathbf{\Omega}$ as having *effective unit rank* since the norm ratio bounded in $\mathcal{O}(1)$. More precisely, the following statement holds:

Lemma S.21 (Ratio of 1 and 2-norms)**.** *Every matrix $X \in \mathbf{\Omega}$ has* effective unit rank *in the following sense:*

$$\frac{\|X\|_1^2}{\|X\|_2^2} \leq 9. \tag{S.88}$$

Proof. From $\|X|_1\|_2 \leq 1$ and the definition of $\mathbf{\Omega}$ it follows that $\|X|_1\|_2 + \frac{1}{2}\|X|_1\|_1 \leq \frac{3}{2}$. Hence $\frac{1}{2}\|X|_1\|_2 + \|X|_1\|_1 \leq 3$. Therefore, we have that $\|X\|_1 \leq \|X|_1\|_1 + \|X|_c\|_1 \leq 3$ from which the assertion follows, because every $X \in \mathbf{\Omega}$ has unit Frobenius norm. ☐

In summary, we want to prove a lower bound on the ℓ_q-norm of the measurement outcomes for trace- and identity annihilating channels with effective unit Kraus rank. The proof uses Mendelson's small ball method. See Ref. [15, Lemma 9] for details of the method as it is stated here, which is a slight generalisation of Tropp's formulation [16] of the original method developed in Refs. [17, 18]. Mendelson's proof strategy requires multiple ingredients. These necessary ingredients will become obvious from the following theorem, which can be found in Ref. [16] and lies at the heart of the small ball method.

Theorem S.22 (Mendelson's small ball method)**.** *Suppose that $\mathcal{A}$ contains m measurements of the form $f_k = \mathrm{Tr}[A_k X]$ where each A_k is an independent copy of a random matrix A. Fix $E \subseteq J(\mathrm{V}_{\mathrm{u,tp,0}})$ and $\xi > 0$ and define*

$$W_m(E; A) := \mathbb{E}\left[\sup_{Z \in E} \mathrm{Tr}\,(ZH)\right], \quad H = \frac{1}{\sqrt{m}} \sum_{k=1}^{m} \epsilon_k A_k, \tag{S.89}$$

$$Q_\xi(E; A) := \inf_{Z \in E} \mathbb{P}\left[|\mathrm{Tr}\,[AZ]| \geq \xi\right], \tag{S.90}$$

where the ϵ_k's are i.i.d. Rademacher random variables, i.e. are uniformly distributed in $\{-1, 1\}$. Then, with probability of at least $1 - \mathrm{e}^{-2t^2}$, where $t \geq 0$,

$$\inf_{Z \in E} \|\mathcal{A}(Z)\|_{\ell_1} \geq \sqrt{m}\left(\xi\sqrt{m} Q_{2\xi}(E; A) - 2W_m(E; A) - \xi t\right).$$

A lower bound of $\|\mathcal{A}(X)\|_{\ell_1}$ thus requires two main ingredients: 1.) a lower bound on the so-called *mean empirical width* $W_m(E; A)$ and 2.) an upper bound on the so-called *marginal tail function* $Q_{2\xi}(E; A)$. We will derive those bounds for $E = \mathbf{\Omega}$ and our measurement map $\mathcal{A}$ at hand.

Bound on the mean empirical width. With a different normalisation the following statement is derived in Ref. [11].

Lemma S.23. *Fix $d = 2^n$ and suppose that the measurement matrices are given by $A_i = \frac{d}{d+1}[J(\mathcal{C}_i) + \mathrm{Id}/d]$ with a gate $\mathcal{C}_i$ chosen uniformly from the Clifford group for all i. Also, assume that $m \geq d^2 \log(d)$. Then*

$$W_m(\mathbf{\Omega}, A) \leq \frac{24}{d+1}\sqrt{\log(d)}. \tag{S.91}$$

The proof is analogous to the one in Refs. [11, 12, 15]. In order to adjust the normalisation we provide a short summary.

Proof. For $Z \in \mathbf{\Omega}$ it holds that

$$(A_i, Z) = \frac{d}{d+1}(J(\mathcal{C}_i), Z). \tag{S.92}$$

The constant shift by the identity matrix does not appear hear since every $Z \in \mathbf{\Omega}$ is trace-less. Thus, we can set $H = \frac{d}{\sqrt{m}(d+1)}\sum_{i=1}^m \epsilon_i J(C_i)$. Applying Hölder's inequality for Schatten norms to the definition of the mean empirical width yields

$$W_m(\mathbf{\Omega}, A) \leq \sup_{Z\in\mathbf{\Omega}} \|Z\|_1 \mathbb{E}\|H\|_\infty \leq 3\,\mathbb{E}\|H\|_\infty, \tag{S.93}$$

where we have used the effective unit rank of Z, Lemma S.21. Also, the ϵ_i's in the definition of H form a Rademacher sequence. The non-commutative Khintchine inequality, see e.g [19, Eq. (5.18)], can be used to bound this sequence

$$\mathbb{E}_{\epsilon_i, C_i}\|H\|_\infty \leq \frac{d}{d+1}\sqrt{\frac{2\log(2d^2)}{m}\mathbb{E}_{C_i}\left\|\sum_{i=1}^m J(C_i)^2\right\|_\infty} \tag{S.94}$$

and $J(C_i)^2 = J(C_i)$ further simplifies the remaining expression. Moreover, $\mathbb{E}[J(C_i)] = \frac{1}{d^2}\mathbb{I}$, $\|J(C_i)\|_\infty = 1$ and a Matrix Chernoff inequality for expectations (with parameter $\theta = 1$), see e.g. [20, Theorem 5.1.1] implies

$$\mathbb{E}_{C_i}\left\|\sum_{i=1}^m J(C_i)\right\|_\infty \leq (\mathrm{e}-1)\,\frac{m}{d^2} + \log(d^2) \leq 4\frac{m}{d^2}, \tag{S.95}$$

where the second inequality follows from the assumption $m \geq d^2\log(d)$. Inserting this bound into Eq. (S.94) yields

$$\mathbb{E}_{\epsilon_i, C_i}\|H\|_\infty \leq \frac{d}{d+1}\sqrt{\frac{8\log(2d^2)}{d^2}} \tag{S.96}$$

and the claim follows from combining this estimate with the bound (S.93) and $\log(2d^2) \leq 4\log(d)$. □

Bound on the marginal tail function. Here, we establish an anti-concentration bound to the marginal tail function. The precise result is summarised in the following statement.

Lemma S.24. *Suppose the random variable $A \in H_d$ is given by $A = \frac{d}{d+1}[J(\mathcal{C}) + \mathrm{Id}/d]$, where $\mathcal{C}$ is a Clifford channel drawn uniformly from the Clifford-group $\mathrm{Cl}(d)$. For $0 \leq \xi \leq \frac{1}{d(d+1)}$ it holds that*

$$Q_\xi(\mathbf{\Omega}, A) \geq \frac{1}{\hat{C}}\left(1 - d^2(d+1)^2\xi^2\right)^2, \tag{S.97}$$

where $\hat{C}$ is the constant from Lemma S.25.

This statement follows from applying the Paley-Zygmund inequality to the non-negative random variable $S_{\mathcal{T}}^2$ defined in Eq. (S.3). For this purpose, we will make use of the bounds on the second and fourth moment of $S_{\mathcal{T}}$ derived in Section B and Section C, respectively. In particular, we establish the following relation between the second and fourth moment of $S_{\mathcal{T}}$. This is one of the technical core result of this work.

Lemma S.25. *Let $\mathcal{T} \in V_{\mathrm{u,tp,0}}$ be a map with $J(\mathcal{T})$ of effective unit rank, i.e. $\|J(\mathcal{T})\|_2^2 \leq c\,\|J(\mathcal{T})\|_1^2$ with some constant $c > 0$, then*

$$\mathbb{E}_{U\sim\mathrm{Haar}(\mathrm{Cl}(d))}[S_{\mathcal{T}}^4] \leq \hat{C}\,\mathbb{E}_{U\sim\mathrm{Haar}(\mathrm{Cl}(d))}[S_{\mathcal{T}}^2]^2 \tag{S.98}$$

for some constant $\hat{C}$ independent of the dimension d.

Proof. Since the Clifford group is a unitary 3-design [21, 22], Corollary S.9 implies

$$\mathbb{E}_{U\sim\mathrm{Haar}(\mathrm{Cl}(d))}[S_{\mathcal{T}}^2] \geq \|J(\mathcal{T})\|_2^2\,. \tag{S.99}$$

Furthermore, the effective unit rank assumption, $\|J(\mathcal{T})\|_1^2 \leq c\,\|J(\mathcal{T})\|_2^2$, together with Lemma S.15 yields for the fourth moment

$$\mathbb{E}_{U\sim\mathrm{Haar}(Cl(d))}[S_{\mathcal{T}}^4] \leq \hat{C}\,\|J(\mathcal{T})\|_2^4 \tag{S.100}$$

for some constant $\hat{C} = cC > 0$ independent of d. Combining these two equations, the statement of the proposition follows. □

Note that with the help of Lemma S.11 one arrives at the same conclusion for the moments of $S_{\mathcal{T}}$ when the average is taken over the unitary group. This reproduces the previous technical core result of Ref. [11].

Proof of Lemma S.24. In the following we always understand by $\mathcal{T}$ the map in $\mathrm{L}(H_d)$ with Choi matrix $T = J(\mathcal{T})$. In terms of the random variable $S_{\mathcal{T}} = d^2\,\mathrm{Tr}[TJ(\mathcal{C})]$, Eq. (S.3), the marginal tail function can be expressed as

$$Q_\xi(\mathbf{\Omega}, A) = \inf_{\mathcal{T}\in\mathbf{\Omega}} \mathbb{P}\left[\frac{|S_{\mathcal{T}}|}{d(d+1)} \geq \xi\right]. \tag{S.101}$$

Here we again used that every $Z \in \mathbf{\Omega}$ is trace-less. Consequently, the shift by the identity matrix in the measurements A_i vanishes. Using Lemma S.25, the theorem follows by a straight-forward application of the Paley-Zygmund inequality,

$$\begin{aligned}&\inf_{T\in\mathbf{\Omega}} \mathbb{P}\left[\frac{1}{d(d+1)}|S_{\mathcal{T}}| \geq \xi\right]\\ &= \inf_{T\in\mathbf{\Omega}} \mathbb{P}\left[\frac{1}{d^2(d+1)^2} S_{\mathcal{T}}^2 \geq \frac{\mathbb{E}[S_{\mathcal{T}}^2]}{d^2(d+1)^2}\,\tilde{\xi}^2\right] \\ &\geq (1-\tilde{\xi}^2)^2 \frac{\mathbb{E}[S_{\mathcal{T}}^2]^2}{\mathbb{E}[S_{\mathcal{T}}^4]} \geq \frac{1}{\hat{C}}(1-\tilde{\xi}^2)^2,\end{aligned} \tag{S.102}$$

where $\hat{C} > 0$ and $\tilde{\xi} = \frac{d(d+1)}{\sqrt{\mathbb{E}[S_{\mathcal{T}}^2]}}\,\xi$ is required to fulfil $\tilde{\xi} \in [0,1]$. According to Corollary S.9 and the normalisation of $T \in \mathbf{\Omega}$ we have $\tilde{\xi} = \frac{d(d+1)\,\xi}{\|T\|_2} = d(d+1)\,\xi$. □

Completing the proof of Lemma S.19 We are finally in position to deliver the proof for the NSP of $\mathcal{A}$. With the bounds on the mean empirical width, Lemma S.23, and the marginal tail function, Lemma S.24, Mendelson's small ball method, Theorem S.22, yields the following lemma:

Lemma S.26. *Suppose that $\mathcal{A}$ contains*

$$m \geq m_0 = c\, d^2 \log(d) \tag{S.103}$$

measurements of the form $f_k = \mathrm{Tr}[A_k X]$ where each $A_k = \frac{d}{d+1}\,[J(\mathcal{C}_i) + \mathrm{Id}\,/d]$ is given by an independent and uniformly random Clifford unitary channel $\mathcal{C}_i$. Fix $\mathbf{\Omega} \subset J(\mathrm{V}_{\mathrm{u,tp},0})$ as defined in Lemma S.20. Then

$$\inf_{Z\in\mathbf{\Omega}} \|\mathcal{A}(Z)\|_{\ell_1} \geq C\frac{m}{d(d+1)} \tag{S.104}$$

with probability at least $1-\mathrm{e}^{-c_f m}$ over the random measurements. The constants $C, c, c_f > 0$ only depend on each other.

Proof. Combining the Lemmas S.22, S.23, and S.24 yields with probability at least $1-\mathrm{e}^{-2t^2}$ that

$$\begin{aligned}&\inf_{\mathcal{Z}\in\mathbf{\Omega}} \|\mathcal{A}(\mathcal{Z})\|_{\ell_1}\\ &\geq \sqrt{m}\left(\frac{\xi\sqrt{m}}{\hat{C}}\left(1-(d(d+1)\xi)^2\right)^2 - \frac{48}{d+1}\sqrt{\log(d)} - \xi t\right)\\ &\geq \frac{\sqrt{m}}{d+1}\left(c_1\frac{\sqrt{m}}{d} - 48\sqrt{\log(d)} - \frac{t}{2d}\right)\end{aligned} \tag{S.105}$$

where we have chosen $\xi = \frac{1}{2d(d+1)}$. The statement follows from the scaling (S.103) of m. □

From Lemma S.26 and Lemma S.20 the assertion of Lemma S.19 directly follows.

E. Sample optimality in the number of channel uses

The compressed sensing recovery guarantees, Theorem 2 and Theorem S.16, focus on the minimal number of AGFs m that are required for the reconstruction of a unital and trace-preserving quantum channel using the reconstruction procedure (3) and (S.79), respectively. This can be regarded as the number of measurement settings. But already the measurement of single fidelities up to some desired additive error will require a certain number of repetitions of some experiment. Therefore, to quantify the total measurement effort a more relevant figure of merit is the minimum number of channel uses M required for taking all the data used in a reconstruction.

We will show that the equivalent algorithms (3) and (S.79) reach an optimal parametric scaling of the required number of channel uses in a simplified measurement setting. To this end, we first combine the direct fidelity estimation protocol of Ref. [23] with our recovery strategy to provide an upper bound on the number of channel uses required for the reconstruction of a unitary gate up to a constant error. Second, following the proof strategy of Ref. [24, Section III], we derive a lower bound on the number of channel uses required by any POVM measurement scheme of AGFs with Clifford gates and any subsequent reconstruction protocol that only relies on these AGFs.

1. *Measurement setting*

In order to obtain an optimality result we consider a measurement setting that is arguably simpler than the one in randomised benchmarking and more basic from a theoretical perspective. We consider a unitary channel $\mathcal{U}$ given by a unitary $U \in \mathrm{U}(d)$ and measurements given by Clifford channels $\mathcal{C}_i$ with $C_i \in \mathrm{Cl}(d)$. Using the identities (S.1) and (S.2) the AGFs $F_{\mathrm{avg}}(\mathcal{C}_i, \mathcal{X})$ are determined by

$$f_i = (J(\mathcal{C}_i), J(\mathcal{X})) = \frac{1}{d^2}|\,\mathrm{Tr}[C_i U]|^2\,. \tag{S.106}$$

In this section, we consider $U/\sqrt{d}$ as a pure state vector in $\mathbb{C}^d \otimes \mathbb{C}^d$, i.e., as the state vector corresponding to the Choi state of the channel $\mathcal{U}$. This state can be prepared by applying the operation U to one half of a maximally entangled state.

2. *An upper bound from direct fidelity estimation*

We will now derive an upper bound on the number of channel uses required in the reconstruction scheme (S.79). We note that our measurement values (S.106) are also fidelities of the quantum state vectors $U/\sqrt{d}$ and $C_i/\sqrt{d}$ and use *direct fidelity estimation* [23] (see also [25]) to estimate these fidelities. Importantly, each $C_i/\sqrt{d}$ is a stabiliser state and we view it as the "target state" in the direct fidelity estimation protocol [23]. Then $C_i/\sqrt{d}$ is a *well-conditioned state* with parameter $\alpha = 1$. One of the main statements of Ref. [23] is that the fidelity f_i can hence be estimated from $\mu \geq \mu_0$ many Pauli measurements, where $\mu_0 \in \mathcal{O}\left(\frac{\log(1/\delta_0)}{\varepsilon_{\mathrm{F}}^2}\right)$. Here, $\delta_0 > 0$ is the maximum failure probability, and $\varepsilon_{\mathrm{F}} > 0$ is the accuracy up to which the fidelity f_i is estimated. This implies that the estimation error is bounded as

$$\varepsilon_{\mathrm{F}} \in \mathcal{O}\left(\frac{\sqrt{\log(1/\delta_0)}}{\sqrt{\mu_0}}\right). \tag{S.107}$$

For our channel reconstruction, we measure $m \in \tilde{\mathcal{O}}(d^2)$ many fidelities, each up to error ε_{F}, see Theorem 2. For a maximum failure probabilities of the single fidelity estimations δ_0 and a desired failure probability δ of all the m estimations it is sufficient to require $\delta \leq m\delta_0$, since $(1-\delta_0)^m \geq 1 - m\delta_0$. Moreover, in order for the reconstruction error (5) to be bounded as

$$\hat{C}\frac{d^2}{\sqrt{m}}\,\|\epsilon\|_{\ell_2} \leq \varepsilon_{\mathrm{rec}}\,, \tag{S.108}$$

where $\|\epsilon\|_{\ell_2} \leq \sqrt{m}\,\varepsilon_{\mathrm{F}}$, we require

$$\hat{C}\frac{d^2}{\sqrt{m}}\,\|\epsilon\|_{\ell_2} \leq C_2 d^2 \frac{\sqrt{\log(m/\delta)}}{\sqrt{\mu_0}} \leq \varepsilon_{\mathrm{rec}}\,. \tag{S.109}$$

Thus, a constant bound $\varepsilon_{\mathrm{rec}}$ of the reconstruction error can be achieved with a number of channel uses M in

$$\mathcal{O}\left(\frac{d^4\log(m/\delta)}{\varepsilon_{\mathrm{rec}}^2}\right) \subset \tilde{\mathcal{O}}\left(\frac{d^4}{\varepsilon_{\mathrm{rec}}^2}\right). \tag{S.110}$$

3. *Information theoretic lower bound on the number of channel uses*

In this section we derive a lower bound on the number of channel uses that holds in a general POVM framework. Up to log-factors, it has the same dimensional scaling as the upper bound (S.110) from direct fidelity estimation.

We extend the arguments of Ref. [24, Section III] to prove a lower bound on the number of channel uses required for QPT of unitary channels from measurement values of the form (S.106). We consider each of these values to be an expectation value in a binary POVM measurement setting given by the unit rank projector $J(\mathcal{C}_i)$ are applied to the Choi state $J(\mathcal{U})$. Then we are in the situation of [24, Section 3], which proves a lower bound for the *minimax risk* – a prominent figure of merit for statistical estimators.

Let us summarise this setting. We denote by $\mathcal{S} \subset H_d$ the set of density matrices and by $\mathcal{M}$ the set of all two-outcome positive-operator-valued measurements (POVMs), each of them given by a projector $\pi \in H_d$. Next, we assume that we measure M copies of an unknown state $\rho \in \mathcal{S}$ in a sequential fashion. By Y_i we denote the binary random variable that is given by choosing the i-th measurement $\pi_i \in \mathcal{M}$ and measuring ρ. These are mapped to an estimate $\hat{\rho}(Y_1, \dots, Y_M) \in H_d$. Any such estimation protocol is specified by the estimator function $\hat{\rho}$ and a set of functions $\{\Pi_i\}_{i\in[M]}$ that correspond to the measurement choices, where $\Pi_i(Y_1, \dots Y_{i-1}) \in \mathcal{M}$, i.e., the i-th measurement choice Π_i only depends on previous measurement outcomes. Let $\varepsilon > 0$ be the maximum trace distance error we like to tolerate between the estimation $\hat{\rho}$ and ρ. Then the *minimax risk* is defined as

$$R^*(M,\varepsilon) := \inf_{\substack{\hat{\rho} \\ \Pi_1,\dots,\Pi_M}} \sup_{\rho\in\mathcal{S}} \mathbb{P}\left[\|\hat{\rho}(Y) - \rho\|_1 > \varepsilon\right], \tag{S.111}$$

where we denote by Y the vector consisting of all random variables Y_i. An estimation protocol $(\hat{\rho}, \{\Pi_i\}_{i\in[M]})$ minimising the minimax risk has the smallest possible worst-case probability over the set of quantum states.

The following theorem provides a lower bound on the minimax risk for the estimation of the Choi matrix of a unitary gate from unit rank measurements.

Theorem S.27 (Lower bound, unit rank measurements)**.** *Fix a set $\mathcal{M}$ of rank-1 measurements. For $\varepsilon > 0$ the minimax risk* (S.111) *of measurements of M copies is bounded as*

$$R^*(M,\varepsilon) \geq 1 - c_1\frac{\log(d)\log(|\mathcal{M}|)}{d^4(1-\varepsilon/2)^2}M - \frac{c_2}{d^2(1-\varepsilon^2)}\,, \tag{S.112}$$

where c_1 and c_2 are absolute constants.

Before providing a proof for this theorem let us work out its consequences. If the measurements project onto Clifford unitaries, we get the following lower bound on the minimax risk.

Corollary S.28 (Lower bound, Clifford group)**.** *Let $\varepsilon > 0$ and consider measurements of the form* (S.106) *given by Clifford group unitaries on M copies. Then the minimax risk* (S.111) *is bounded as*

$$R^*(M,\varepsilon) \geq 1 - c_3\frac{\log(d)^3}{d^4(1-\varepsilon/2)^2}M - \frac{c_2}{d^2(1-\varepsilon^2)}\,, \tag{S.113}$$

where c_3 and c_2 are absolute constants.

Proof. The cardinality of the n-qubit Clifford group ($d = 2^n$) is bounded as

$$|\operatorname{Cl}(d)| = 2^{n^2+2n}\prod_{j=1}^{n}(4^j - 1) < 2^{2n^2+4n} \tag{S.114}$$

[26]. This implies that in case of our Clifford group measurements we have $\log(|\mathcal{M}|) < 2\log(d)^2 + 4\log(d)$. ☐

In every meaningful measurement and reconstruction scheme the minimax risk needs to be small. The corollary implies that, in the case of Cliffords, the number of copies M need to scale with the dimension as

$$M \in \Omega\left(\frac{d^4}{\log(d)^3}\right), \tag{S.115}$$

where we have assumed $\varepsilon > 0$ to be small. This establishes a lower bound on the number of channel uses that every POVM measurement and reconstruction scheme requires for a guaranteed successful recovery of unitary channels from AGFs with respect to Clifford unitaries.

From the argument as it is presented here it is not possible to extract the optimal parametric dependence of the number of channel uses M on the desired reconstruction error ε. For quantum state tomography such bounds were derived in Ref. [27] by extending the argument of Ref. [24] and constructing different ε-packing nets. By adapting the ε-packing net constructions of Ref. [27] to unitary gates one might be able to derive a optimal parametric dependence of M on ε. But it is not obvious how one can incorporate the restriction of the measurements to unit rank in the argument of Ref. [27]. We leave this task to future work.

In the remainder of this section we prove Theorem S.27. The proof proceeds in two steps. At first we derive a more general bound on the minimax risk, Lemma S.29, that follows mainly from combining Fano's inequality with the data processing inequality, see e.g. [28]. This is a slight generalization of Lemma 1 of Ref. [24] adjusted to the situation where the outcome probabilities of the POVM measurements do not necessarily concentrate around the value $1/2$. Lemma S.29 assumes the existence of an ε-packing net for the set of unitary gates whose measurement outcomes are in a small interval to establish a lower bound on the minimax risk. Hence, in order to complete the proof, we have to establish the existence of a suitable packing net, Lemma S.33, in a second step. Combining the general bound of Lemma S.29 and the existence of the packing net of Lemma S.33, the proof of Theorem S.27 follows.

We begin with the general information theoretic bound on the minimax risk.

Lemma S.29 (Lower bound to the minimax risk)**.** *Let $\varepsilon > 0$ and $0 < \alpha < \beta \leq 1/2$. Assume that there are states $\rho_1, \dots, \rho_s \in \mathrm{Pos}_D$ and orthogonal projectors $\pi_1, \dots, \pi_n \in \mathrm{Pos}_D$ such that*

$$\|\rho_i - \rho_j\|_1 \geq \varepsilon \tag{S.116}$$

$$\mathrm{Tr}[\pi_k \rho_i] \in [\alpha, \beta] \tag{S.117}$$

for all $i \neq j \in [s]$ and $k \in [n]$. Then the minimax risk (S.111) of M single measurements is bounded as

$$R^*(M, \varepsilon) \geq 1 - \frac{M(h(\beta) - h(\alpha)) + 1}{\log(s)}\,, \tag{S.118}$$

where h denotes the binary entropy.

Proof. We start by following the proof of [24, Lemma 1]: Let X be the random variable uniformly distributed over $[s]$ and let $Y_1, \dots, Y_M$ be the random variables describing the M single POVM measurements performed on ρ_X. Consider any estimator $\hat\rho$ of the state ρ_X from the measurements Y and define

$$\hat X(Y) \coloneqq \operatorname*{arg\,min}_{i \in [s]} \|\hat\rho(Y) - \rho_i\|_1\,. \tag{S.119}$$

Then, for all $i \in [s]$,

$$\mathbb{P}[\|\hat\rho(Y) - \rho_i\|_1 \geq \varepsilon] \geq \mathbb{P}[\hat X(Y) \neq X]. \tag{S.120}$$

Following Ref. [24], we combine Fano's inequality and the data processing inequality for the mutual information $I(X;Z) = H(X) - H(X|Z)$, where H denotes the entropy and conditional entropy, to obtain

$$\mathbb{P}[\hat X(Y) \neq X] \geq \frac{H(X|\hat X(Y)) - 1}{\log(s)} \tag{S.121}$$

$$\geq 1 - \frac{I(X;Y) + 1}{\log(s)}\,. \tag{S.122}$$

Now we start deviating from Ref. [24]. We use that $I(X;Y) = I(Y;X)$, the chain rule, and the definition of the conditional entropy to obtain

$$\mathbb{P}[\hat X(Y) \neq X] \tag{S.123}$$

$$\geq 1 - \frac{H(Y) - H(Y|X) + 1}{\log(s)} \tag{S.124}$$

$$= 1 - \frac{1}{\log(s)}\Bigg(\sum_{j=1}^{M}\Big\{H(Y_j|Y_{j-1}, \dots, Y_1) \tag{S.125}$$

$$- \frac{1}{s}\sum_{i=1}^{s} H(Y_j|Y_{j-1}, \dots, Y_1, X = i)\Big\} + 1\Bigg). \tag{S.126}$$

Now we use that

$$H(Y_j|Y_{j-1}, \dots, Y_1, X = i) \geq h(\alpha) \tag{S.127}$$

and

$$H(Y_j|Y_{j-1}, \dots, Y_1) \leq h(\beta)\,, \tag{S.128}$$

where h is the binary entropy, to arrive at

$$\mathbb{P}[\hat X(Y) \neq X] \geq 1 - \frac{M(h(\beta) - h(\alpha)) + 1}{\log(s)} \tag{S.129}$$

$$\geq 1 - \frac{M(h(\beta) - h(\alpha)) + 1}{\log(s)}. \tag{S.130}$$

□

To apply Lemma S.29 we need to proof the existence of an ε-packing net $\{\rho_i\}_{i=1}^s$ consisting of unitary quantum gates with the properties (S.116) and (S.117). The construction of such a suitable ε-packing net will use the fact that the modulus of the trace of a Haar random unitary matrix is a sub-Gaussian random variable. This can be viewed as a non-asymptotic version of a classic result by Diaconis and Shahshahani [29]: the trace of a Haar random unitary matrix in $\mathrm{U}(d)$ is a complex Gaussian random variable in the limit of infinitely large dimensions d.

The trace of Haar random unitaries is sub-Gaussian. The statement follows from the fact that the moments of the modulus of the trace of a Haar random unitary are dominated by the moments of a Gaussian variable.

Proposition S.30. *For all $d, k \in \mathbb{Z}_+$*

$$\mathbb{E}_{U \sim \mathrm{Haar}(\mathrm{U}(d))}\left[|\,\mathrm{Tr}[U]|^{2k}\right] \leq k!, \tag{S.131}$$

with equality if $k \leq d$.

Proof. Denote by $S \coloneqq |\,\mathrm{Tr}(U)|^2$ the random variable with $U \in \mathrm{U}(d)$ drawn from the Haar measure. Let $\{\,|n\rangle\}_{n=1}^{d^k}$ be an orthonormal basis of $(\mathbb{C}^d)^{\otimes k}$. The k-th moment of S is given by

$$\mathbb{E}[S^k] = \sum_{n,m=1}^{d^k} \langle n|\, U^{\otimes k}\, |n\rangle\, \langle m|\, (U^\dagger)^{\otimes k}\, |m\rangle\,. \tag{S.132}$$

Applying Theorem S.2, we get

$$\mathbb{E}[S^k] = \frac{1}{k!}\sum_{n,m=1}^{d^k}\sum_{\tau\in S_k}\sum_{\lambda\vdash k, l(\lambda)\le d}\frac{d_\lambda}{D_\lambda} \tag{S.133}$$

$$\times \langle m|\,\pi^d_{S_k}(\tau)\,|n\rangle\,\langle n|\,\pi^d_{S_k}(\tau^{-1})P_\lambda\,|m\rangle \tag{S.134}$$

$$= \frac{1}{k!}\sum_{\tau\in S_k}\sum_{\lambda\vdash k, l(\lambda)\le d}\frac{d_\lambda}{D_\lambda}\operatorname{Tr}(\pi^d_{S_k}(\tau)\pi^d_{S_k}(\tau^{-1})P_\lambda) \tag{S.135}$$

$$= \sum_{\lambda\vdash k, l(\lambda)\le d}\frac{d_\lambda}{D_\lambda}\operatorname{Tr}(P_\lambda). \tag{S.136}$$

Since $\operatorname{Tr}(P_\lambda) = d_\lambda D_\lambda$, we conclude

$$\mathbb{E}[S^k] = \sum_{\lambda\vdash k, l(\lambda)\le d} d_\lambda^2 \le \sum_{\lambda\vdash k} d_\lambda^2 = k!\,. \tag{S.137}$$

The last equality can be seen from the orthogonality relation of the characters of the symmetric group, see e. g. Ref. [7, Chapter 2] for more details. Note that the second inequality is saturated in the case where $k \le d$ since in this case the restriction $l(\lambda) \le d$ is automatically fulfilled. □

As a simple implication of the previous lemma is that the random variable $S = |\operatorname{Tr}(U)|^2$ has subexponential tail decay.

Lemma S.31. *Let S be a real-valued random variable that obeys $\mathbb{E}\left[|S|^k\right] \le k!$ for all $k \in \mathbb{N}$. Then, the right tail of X decays at least subexponentially. For any $t \ge 0$,*

$$\mathbb{P}\left[S \ge t\right] \le \mathrm{e}^{-\kappa t+2},$$

with $\kappa = 1 - \frac{1}{2\mathrm{e}}$.

This is a consequence of a standard result in probability theory that can be found in many textbooks, e.g. [19] and [14, Section 7.2]. We present a short proof here in order to be self-contained.

Proof. We use Markov's inequality, Proposition S.30, and Stirling's bound $k! \le \mathrm{e}\sqrt{k}\,k^k\,\mathrm{e}^{-k}$ to obtain for any $k \in \mathbb{N}$

$$\mathbb{P}[S \ge k] \le \frac{\mathbb{E}[|S|^k]}{k^k} \le \frac{k!}{k^k} \le \mathrm{e}\sqrt{k}\mathrm{e}^{-k}. \tag{S.138}$$

In order to prove the tail bound, we choose $t \ge 0$ arbitrary and let k be the largest integer that is smaller or equal to t ($k = \lfloor t \rfloor$). Then

$$\Pr\left[S \ge t\right] \le \Pr\left[S \ge k\right] \le \mathrm{e}\sqrt{k}\mathrm{e}^{-k} \le \mathrm{e}^{-\kappa k+1} \le \mathrm{e}^{-\kappa t+1+\kappa}.$$

Here, we have used $\sqrt{k}\mathrm{e}^{-k} \le \mathrm{e}^{-\kappa k}$ and $t \le k+1$. □

Random variables with subgaussian tail decay – *subgaussian random variables* – are closely related to random variables with subexponential tail decay: X is subgaussian if and only if X^2 is subexponential.

Thus, Proposition S.30 highlights that the trace of a Haar-random unitary is a subgaussian random variable. This is the aforementioned generalization of the classical result by Diaconis and Shashahani.

A packing net with concentrated measurements. The proof of existence of an ε-packing net to apply Lemma S.29 uses a probabilistic argument as in Ref. [24]. Here, the strategy is the following: We assume we are already given an ε-packing net of a size $s-1$ that satisfies the desired concentration condition (S.117). We then show that a Haar random unitary gate also fulfils the concentration condition and is ε-separated from the rest of the net with strictly positive probability. Consequently, if one can be lucky to randomly arrive at a suitable ε-packing net of size s in this way then it must also exist.

We start by deriving an anti-concentration result for the Choi matrix $J(\mathcal{U})$ of a unitary channel given by a Haar random unitary U in $\mathrm{U}(d)$.

Lemma S.32. *Let $\mathcal{V}$ be a unitary gate. For all $\varepsilon > 0$*

$$\mathbb{P}_{U\sim \mathrm{Haar}(\mathrm{U}(d))}[\|J(\mathcal{U}) - J(\mathcal{V})\|_1 \le \varepsilon] \le \mathrm{e}^{-\kappa d^2(1-\varepsilon/2)^2+2} \tag{S.139}$$

with $\kappa > 0$ being the constant from Lemma S.31.

Proof. Due to the unitary invariance of the trace-norm and the Haar measure, it suffice to show the statement for $\mathcal{V} = \mathrm{Id}$. For a unitary channel with Choi-matrix $J(\mathcal{U}) = d^{-1}\operatorname{vec}(U)\operatorname{vec}(U^\dagger)^t$ and Kraus-operator $U \in \mathrm{U}(d)$ we have

$$\|J(\mathcal{U}) - J(\mathrm{Id})\|_1 = 2\sqrt{1 - \frac{1}{d^2}|\operatorname{Tr}(U)|^2} \ge 2\left(1 - \frac{1}{d}|\operatorname{Tr}(U)|\right). \tag{S.140}$$

For the first equation we calculate the set eigenvalues of $J(\mathcal{U}) - J(\mathrm{Id})$, which is $\{\pm\sqrt{1-d^{-2}|\operatorname{Tr}(U)|^2}\}$. Introducing the random variable $S_U := |\operatorname{Tr}(U)|^2$, we can rewrite the probability as

$$\mathbb{P}[\|J(\mathcal{U}) - J(\mathrm{Id})\|_1 \le \varepsilon] \le \mathbb{P}\left[2\left(1 - \frac{1}{d}\sqrt{S_U}\right) \le \varepsilon\right] \tag{S.141}$$

$$= \mathbb{P}\left[S_U \ge d^2\left(1 - \frac{\varepsilon}{2}\right)^2\right]. \tag{S.142}$$

From Lemma S.31 we know that

$$\mathbb{P}\left[S_U \ge d^2\left(1 - \frac{\varepsilon}{2}\right)^2\right] \le \mathrm{e}^{-\kappa d^2(1-\varepsilon/2)^2+2} \tag{S.143}$$

from which the assertion follows. □

The anti-concentration result of Lemma S.32 implies the existence of a large ε-packing net $\mathcal{N}_\varepsilon$ of unitary quantum channels. The desired concentration of the measurement outcomes can be established using Lemma S.31. In summary we arrive at the following assertion:

Lemma S.33 (Packing net with concentrated measurements)**.** *Let $0 < \varepsilon < 1/2$, $\kappa = 1 - \frac{1}{2\mathrm{e}}$, and $C_1, \dots, C_K \in \mathrm{U}(d)$. Then, for any number $s < \frac{1}{2}\mathrm{e}^{\kappa(1-\varepsilon/2)^2 d^2 - 2}$, there exist $U_1, \dots, U_s \in \mathrm{U}(d)$ such that for all $i, j \in [s]$ with $i \ne j$ and for all $k \in [K]$*

$$\|J(\mathcal{U}_i) - J(\mathcal{U}_j)\|_1 \ge \varepsilon\,, \tag{S.144}$$

$$\frac{1}{d^2}|\operatorname{Tr}[C_k^\dagger U_i]|^2 \le \frac{\log(2K)+2}{\kappa d^2}\,. \tag{S.145}$$

Proof. As outlined above the existence of the described ε-packing net follows inductively from the fact that if one adds a Haar random unitary gate $\mathcal{U}$ to an ε-packing $\tilde{\mathcal{N}}_\varepsilon$ of size $s-1$ that already fulfils all requirements of the lemma the resulting set $\tilde{\mathcal{N}}_\varepsilon \cup \{\mathcal{U}\}$ has still a strictly positive probability to be an ε-packing net with the desired concentration property (S.145).

We start with bounding the probability that the resulting set $\tilde{\mathcal{N}}_\varepsilon \cup \{\mathcal{U}\}$ fails to be an ε-packing net. Let us denote the probability that a Haar random $\mathcal{U}$ is not ε-separated from $\tilde{\mathcal{N}}_\varepsilon$ by $\bar{p}_\varepsilon$. In other words, $\bar{p}_\varepsilon$ is the probability that there exists $\mathcal{V} \in \tilde{\mathcal{N}}_\varepsilon$ with

$$\|J(\mathcal{U}) - J(\mathcal{V})\|_1 \leq \varepsilon. \tag{S.146}$$

Taking the union bound for all $\mathcal{V} \in \tilde{\mathcal{N}}_\varepsilon$, Lemma S.32 implies that

$$\bar{p}_\varepsilon \leq s\mathrm{e}^{-\kappa d^2(1-\epsilon/2)^2+2} \tag{S.147}$$

with $\kappa = 1 - \frac{1}{2\mathrm{e}}$. Thus, for $s < \frac{1}{2}\mathrm{e}^{-\kappa d^2(1-\epsilon/2)^2+2}$ we ensure that $\bar{p}_\varepsilon < \frac{1}{2}$.

We now also have to upper bound the probability $\bar{p}_c$ of $\mathcal{U}$ not having a concentration property

$$\frac{1}{d^2}|\operatorname{Tr}[C_k^\dagger U_i]|^2 \leq \beta \tag{S.148}$$

with respect to K different unitaries $C_1, \dots, C_K$.

Using the unitary invariance of the Haar measure and taking the union bound, the tail-bound for the squared modulus of the trace of a Haar random unitary, Lemma S.31, yields

$$\bar{p}_c \leq K\mathrm{e}^{-\kappa\beta d^2+2} \tag{S.149}$$

for $\beta \geq 2$. In order for $\bar{p}_c$ to be at most $1/2$, we need that

$$\beta \geq \frac{\log(2K)+2}{\kappa d^2}. \tag{S.150}$$

In summary, we have established that $\bar{p}_\varepsilon + \bar{p}_c < 1$ as long as $s < \frac{1}{2}\mathrm{e}^{-\kappa d^2(1-\epsilon/2)^2+2}$ and the achievable concentration is $\beta \geq (\log(2K)+2)/(\kappa d^2)$. Hence, in this parameter regime there always exist at least one additional unitary gate extending the ε-packing net. Inductively this proves the existence assertion of the lemma. □

Having established a suitable ε-packing net, we can now apply Lemma S.29 to derive the lower bound on the minimax-risk for the recovery of unitary gates from unit rank measurements of Theorem S.27, the main result of this section.

Proof of Theorem S.27. We will apply Lemma S.29 with $\alpha = 0$ and

$$\beta = \frac{\log(2|\mathcal{M}|)+2}{\kappa d^2} \tag{S.151}$$

and we use that $h(\beta) \leq 2\beta\log(1/\beta)$ for $\beta \leq 1/2$. Combining the Lemmas S.29 and S.33 we obtain

$$R^*(M,\varepsilon) \geq 1 - \frac{Mh(c/d^2)+1}{(\kappa(1-\varepsilon/2)^2d^2+2)/\log(2)-2} \tag{S.152}$$

$$\geq 1 - \frac{2\frac{\log(2|\mathcal{M}|)+2}{\kappa d^2}\log\left(\frac{\log(2|\mathcal{M}|)+2}{\kappa d^2}\right)M+1}{d^2(\kappa(1-\varepsilon/2)^2d^2+2)/\log(2)-2}, \tag{S.153}$$

where, in Lemma S.33 we have chosen s to be the strict upper bound minus one. Finally, we simplify the bound by choosing large enough constants c_1 and c_2. □

F. Expansion of quantum channels in average gate fidelities

In this section, we give a instructive proof of the result of [30] that the linear span of the *unital* channels coincides with the linear span of the *unitary* ones, even if one restricts to the unitaries from a unitary 2-design. We also link this finding to AGFs. On the way, we establish the simple formula of Proposition 1 that allows for the reconstruction of unital and trace-preserving maps from measured AGFs with respect to a arbitrary unitary 2-design, e.g. Clifford gates.

In Lemma S.8 we derived an explicit expression for the second moment of the random variable $S_{\mathcal{T}} = d^2(\mathcal{T},\mathcal{U})$. For $\mathcal{T} \in \mathrm{L}_{\overline{\mathrm{u,tp}}}$, the linear hull of unital and trace-preserving maps, and $\mathcal{U}$ uniformly drawn from a unitary 2-design the expression in fact indicates that a unitary 2-design constitutes a Parseval frame for $\mathrm{L}_{\overline{\mathrm{u,tp}}}$. More abstractly, this observation stems from the general fact that irreducible unitary representations form Parseval frames on the space of endomorphisms of their representation space. For this reason it is instructive, to derive the connection explicitly in the 'natural' representation-theoretic language. We begin with formalising the connection between irreducible representations and Parseval frames.

Lemma S.34 (Irreps form a Parseval frame). *Let $R: G \to \mathrm{L}(V)$ be an irreducible unitary representation of a group G. Then the set $\{\sqrt{\dim V}\, R(g)\}_{g\in G}$ forms a Parseval frame for the space $\mathrm{L}(V)$ equipped with the Hibert-Schmidt inner product $A, B \mapsto \operatorname{Tr}[A^\dagger B]$, in the sense that*

$$T_G(A) := \dim(V)\int_G R(g)\operatorname{Tr}[R(g)^\dagger A]\,d\mu(g) = A \tag{S.154}$$

for all $A \in \mathrm{L}(V)$.

Proof. Since $\mathrm{L}(V)$ is generated as an algebra by $\{R(g)\}_{g\in G}$ (see e.g. [7, Proposition 3.29]), it suffices to show the statement for $A = R(g)$ with $g \in G$. Due to the invariance of the Haar measure, the map T_G is covariant in the sense that $T_G(R(g)B) = R(g)T_G(B)$ for all $B \in \mathrm{L}(V)$. In particular, for $B = \mathrm{Id}$, we thus get $T_G(R(g)\,\mathrm{Id}) = R(g)T_G(\mathrm{Id})$. With $\chi(g) = \operatorname{Tr} R(g)$ the character of the representation, we have

$$T_G(\mathrm{Id}) = \dim(V)\int_G R(g)\bar{\chi}(g)\,d\mu(g) = \mathrm{Id} \tag{S.155}$$

from the well-known expression for projection onto a representation space in terms of the character, see e.g. Ref. [7, Chapter 2.4]. Thus, we have established that $S_R(R(g)) = R(g)$ for all $g \in G$. □

Applying this lemma to unitary channels, we can derive the following expression for the orthogonal projection onto the linear hull of unital and trace-preserving maps.

Theorem S.35. *Let $\{\mathcal{U}_k\}_{k=1}^N$ be a unitary 2-design. The orthogonal projection onto the linear hull of unital and trace-preserving maps $\mathrm{L}_{\overline{\text{u,tp}}}(H_d)$ is give by*

$$P_{\overline{\text{u,tp}}}(\mathcal{X}) = \frac{1}{N}\sum_{k=1}^{N} c_{\mathcal{U}_k}(\mathcal{X})\,\mathcal{U}_k \tag{S.156}$$

with coefficients

$$c_{\mathcal{U}}(\mathcal{X}) = C F_{avg}(\mathcal{U},\mathcal{X}) - \frac{1}{d}\left(\frac{C}{d}-1\right)\mathrm{Tr}(\mathcal{X}(\mathrm{Id}))\,, \tag{S.157}$$

where $C \coloneqq d(d+1)(d^2-1)$.

Proof. Throughout the proof, we denote the unitary channel representing the unitary $U \in U(d)$ on space of Hermitian operators H_d by $\mathcal{U} : \rho \mapsto U\rho U^\dagger$. The vector space H_d is a direct sum of the space $\mathcal{K}_0$ of trace-less hermitian matrices, and of $\mathcal{K}_1 = \{z\,\mathrm{Id}\}_{z\in\mathbb{C}}$. The group of unitary channels acts trivially on $\mathcal{K}_1$, and irreducibly on $\mathcal{K}_0$. In particular, $\mathcal{U}$ is "block-diagonal" $\mathcal{U} = \mathcal{U}_0 \oplus 1$ with respect to this decomposition, where $\mathcal{U}_0 \in \mathrm{L}(\mathcal{K}_0)$ is the irreducible (d^2-1)-dimensional block. More generally, the projection of a map $\mathcal{X}$ onto the linear hull of unital and trace-preserving maps $\mathrm{L}_{\overline{\text{u,tp}}}(H_d)$ is of the form $\mathcal{X}_0 \oplus x_1$. The map $\mathcal{X}_0 \oplus x_1$ is trace-preserving and unital if and only if $x_1 = \mathrm{Tr}(\mathcal{X}(\mathrm{Id}/d)) = 1$. For the map $\mathcal{X} \in \mathrm{L}(H_d)$ we have

$$\mathrm{Tr}[\mathcal{U}^\dagger\mathcal{X}] = \mathrm{Tr}[\mathcal{U}_0^\dagger\mathcal{X}_0] + x_1. \tag{S.158}$$

Using this formula, Lemma S.34 for the choice $V = \mathcal{K}_0$, and the fact that a group integral over a non-trivial irrep vanishes [31] , we find

$$\begin{aligned}
&(d^2-1)\int_{U(d)} \mathcal{U}\ \mathrm{Tr}\left[\mathcal{U}^\dagger\mathcal{X}\right] d\mu(U)\\
&= (d^2-1)\int_{U(d)} (\mathcal{U}_0\oplus 1)(\mathrm{Tr}[\mathcal{U}_0^\dagger\mathcal{X}_0]+x_1)\,d\mu(U)\\
&= (d^2-1)\int_{U(d)} \mathcal{U}_0(\mathrm{Tr}[\mathcal{U}_0^\dagger\mathcal{X}_0]+x_1)\,d\mu(U)\\
&\quad\oplus (d^2-1)\int_{U(d)} (\mathrm{Tr}[\mathcal{U}_0^\dagger\mathcal{X}_0]+x_1)\,d\mu(U)\\
&= \mathcal{X}_0 \oplus (d^2-1)x_1.
\end{aligned} \tag{S.159}$$

Hence, for $\mathcal{X} \in \mathrm{L}_{\overline{\text{u,tp}}}(H_d)$ we obtain the completeness relation

$$\begin{aligned}
\int_{U(d)} \mathcal{U}\left((d^2-1)\,\mathrm{Tr}[\mathcal{U}^\dagger\mathcal{X}] + \frac{2-d^2}{d}\,\mathrm{Tr}[\mathcal{X}(\mathrm{Id})]\right) d\mu(U)&\\
= \mathcal{X}.&
\end{aligned} \tag{S.160}$$

For $\mathcal{X}$ in the ortho-complement of $\mathrm{L}_{\overline{\text{u,tp}}}(H_d)$ the left hand side of Eq. (S.160) vanishes. The expression, thus, defines the orthogonal projection $P_{\overline{\text{u,tp}}}$ onto $\mathrm{L}_{\overline{\text{u,tp}}}$. The projection can be re-expressed in terms of the AGF. With the help of Eqs. (S.1, S.2),

$$\begin{aligned}
\mathrm{Tr}[\mathcal{U}^\dagger\mathcal{X}] &= (\mathcal{L}(\mathcal{U}),\mathcal{L}(\mathcal{X})) = d^2(\mathcal{U},\mathcal{X})\\
&= d(d+1)F_{\text{avg}}(\mathcal{U},\mathcal{X}) - \mathrm{Tr}(\mathcal{X}(\mathrm{Id})).
\end{aligned} \tag{S.161}$$

Hence,

$$P_{\overline{\text{u,tp}}}(\mathcal{X}) = \int_{U(d)} c_{\mathcal{U}}(\mathcal{X})\,\mathcal{U}\,d\mu(U), \tag{S.162}$$

with expansion coefficients

$$\begin{aligned}
c_{\mathcal{U}}(\mathcal{X}) &= d(d+1)(d^2-1)F_{\text{avg}}(\mathcal{U},\mathcal{X})\\
&\quad - \frac{1}{d}\left((d+1)(d^2-1)-1\right)\mathrm{Tr}(\mathcal{X}(\mathrm{Id}))\\
&= C F_{\text{avg}}(\mathcal{U},\mathcal{X}) - \frac{1}{d}\left(\frac{C}{d}-1\right)\mathrm{Tr}(\mathcal{X}(\mathrm{Id})).
\end{aligned}$$

Since the integrand in Eq. (S.162) is linear in $U^{\otimes 2}\otimes \bar{U}^{\otimes 2}$, the completeness relation continues to hold if the Haar integral is replaced by the average

$$\frac{1}{N}\sum_{k=1}^{N} c_{\mathcal{U}_k}(\mathcal{X})\,\mathcal{U}_k = P_{\overline{\text{u,tp}}}(\mathcal{X}) \tag{S.163}$$

over any unitary 2-design $\{\mathcal{U}_k\}_{k=1}^N$. □

In the proof, we have used that linear hull of the unital and trace-preserving maps $\mathrm{L}_{\overline{\text{u,tp}}}$ is given by the space of block diagonal matrices $\mathrm{L}(\mathcal{K}_0)\oplus\mathrm{L}(\mathcal{K}_1)$. If $\mathcal{X}$ is not unital and trace-preserving, the image $\mathcal{X}_{\overline{\text{u,tp}}}$ will thus be equal to $\mathcal{X}$, with the off-diagonal blocks set to zero. In particular, the two-norm deviation of a map $\mathcal{X}$ from its projection onto $\mathrm{L}_{\overline{\text{u,tp}}}$ is given by

$$\begin{aligned}
\|\mathcal{X} - P_{\overline{\text{u,tp}}}(\mathcal{X})\|^2 = \frac{1}{d^3}\Big(&\|\mathcal{X}(\mathrm{Id})\|_2^2\\
&+\|\mathcal{X}^\dagger(\mathrm{Id})\|_2^2 - \frac{2}{d}\,\mathrm{Tr}\left(\mathcal{X}(\mathrm{Id})\right)^2\Big).
\end{aligned} \tag{S.164}$$

Based on the arguments used to establish Theorem S.35, we can derive the following variant, which includes a converse statement.

Theorem S.36 (Informational completeness and unitary designs)**.** *Let $\{\mathcal{U}_k\}_{k=1}^N$ be a set of unitary channels. Then the following are equivalent:*

(i) Every unital and trace-preserving map $\mathcal{X}$ can be written as an affine combination $\mathcal{X} = \frac{1}{N}\sum_{k=1}^N c_k(\mathcal{X})\mathcal{U}_k$ of the $\mathcal{U}_k$, with coefficients given by $c_k(\mathcal{X}) = CF_{avg}(\mathcal{U}_k,\mathcal{X}) - \frac{C}{d} + 1$, where $C = d(d+1)(d^2-1)$.

(ii) The set $\{U_k\}_{k=1}^N$ forms a unitary 2-design.

Proof. To show that (ii) implies (i) we apply Theorem S.35. From Eq. (S.160) we can read of that

$$\frac{1}{N}\sum_{k=1}^{N} c_k(\mathcal{X}) = \mathrm{Tr}[\mathcal{X}(\mathrm{Id}/d)] = 1. \tag{S.165}$$

Thus, the linear expansion of $\mathcal{X}$ in terms of the unitary 2-design is affine.

It remains to establish the converse statement. Let $\{\mathcal{U}_k\}_{k=1}^N$ be a set of unitary channels fulfilling

$$\frac{1}{N}\sum_{k=1}^{N}\mathcal{U}_k\left((d^2-1)\,\mathrm{Tr}[\mathcal{U}_k^\dagger\mathcal{X}] + 2 - d^2\right) = \mathcal{X} \tag{S.166}$$

for all $\mathcal{X} \in \mathrm{L}_{\mathrm{u,tp}}(H_d)$.

A handy criterion for verifying that $\{\mathcal{U}_k\}_{k=1}^N$ is a unitary 2-design can be formulated in terms of its frame potential

$$P = \frac{1}{N^2}\sum_{k,k'=1}^{N} |\,\mathrm{Tr}(U_k^\dagger U_{k'})|^4, \tag{S.167}$$

where again U_k is the unitary matrix defining the unitary channel $\mathcal{U}_k$. A set of unitary gates is a unitary 2-design if and only if $P = 2$ [32, Theorem 2]. In fact, Eq. (S.166) allows to calculate the frame potential as follows.

Inserting $\mathcal{X} = 0 \oplus 1$ (the depolarising channel), we find that

$$\frac{1}{N}\sum_{k=1}^{N}\mathcal{U}_k = 0 \oplus 1. \tag{S.168}$$

Note that this implies that the set $\{\mathcal{U}_k\}_{k=1}^N$ constitutes a unitary 1-design. Therefore, Eq. (S.166) takes the form

$$\frac{1}{N}\sum_{k=1}^{N}\mathcal{U}_k(d^2-1)\,\mathrm{Tr}[\mathcal{U}_k^\dagger\mathcal{X}] + 0 \oplus (2-d^2) = \mathcal{X} \tag{S.169}$$

for all $\mathcal{X} \in \mathrm{L}_{\mathrm{u,tp}}(H_d)$. Let the left hand side of Eq. (S.169) define a linear operator $F : \mathcal{X} \mapsto F(\mathcal{X})$. Then Eq. (S.169) implies

$$\frac{1}{N}\sum_{k'=1}^{N}\mathrm{Tr}[\mathcal{U}_{k'}^\dagger F(\mathcal{U}_{k'})] \tag{S.170}$$

$$= \frac{d^2-1}{N^2}\sum_{k,k'=1}^{N}|\,\mathrm{Tr}(U_{k'}^\dagger U_k)|^4 + 2 - d^2 \tag{S.171}$$

$$= d^2 \tag{S.172}$$

and hence

$$\frac{1}{N^2}\sum_{k,k'=1}^{N}|\,\mathrm{Tr}(U_{k'}^\dagger U_k)|^4 = 2. \tag{S.173}$$

This completes the proof. □

Note that for quantum channels, the affine expansion is *almost* convex in the sense that $c_k(\mathcal{X}) \geq 2 - d^2/N \geq -1/d^2$.

G. A new interpretation for the unitarity

In this section, we provide a proof for Theorem 3 and elaborate on its implications. The proof is most naturally phrased by decomposing the linear hull of unital and trace preserving maps $\mathrm{L}_{\overline{\mathrm{u,tp}}}$ into endomorphism acting on the spaces that carry irreducible representations of the unitary channels. In the proof of Theorem S.35 we have explicitly seen that the projection of any map $\mathcal{X}$ onto $\mathrm{L}_{\overline{\mathrm{u,tp}}}$ has the block-diagonal structure:

$$P_{\overline{\mathrm{u,tp}}}(\mathcal{X}) = \mathcal{X}_0 \oplus x_1,$$

where $x_1 = \mathrm{Tr}\left(\mathcal{X}\left(\mathrm{Id}/d\right)\right)$. For channels that are already unital and trace preserving, this projection acts as the identity and $x_1 = 1$. Particular examples of this class are unitary channels $\mathcal{U} = \mathcal{U}_0 \oplus 1$ and the depolarizing channel $\mathcal{D} = \mathbb{O} \oplus 1$ acting as $\mathcal{D}(X) = \frac{\mathrm{Tr}(X)}{d}\,\mathrm{Id}$ on $X \in H_d$. Unitary channels are also special in the sense that they are normalised with respect to the inner products defined in Eqs. (S.1), (S.2) and (S.161):

$$d^2 = \mathrm{Tr}\left[\mathcal{U}^\dagger\mathcal{U}\right] = (\mathcal{L}(\mathcal{U}), \mathcal{L}(\mathcal{U})) = d^2\,(\mathcal{U},\mathcal{U})\,.$$

In fact, unitary channels are the only maps with this property (provided that we also adhere to our convention of normalizing maps with respect to the trace-norm of the Choi matrix). Combining this feature with the "block diagonal" structure of unitary channels yields

$$d^2 = \mathrm{Tr}\left[\mathcal{U}^\dagger\mathcal{U}\right] = \mathrm{Tr}\left[\mathcal{U}_0^\dagger \oplus 1\,\mathcal{U}_0 \oplus 1\right] = 1 + \mathrm{Tr}\left[\mathcal{U}_0^\dagger\mathcal{U}_0\right].$$

This computation implies that a map $\mathcal{X}$ is unitary if and only if

$$u(\mathcal{X}) := \frac{\mathrm{Tr}\left[\mathcal{X}_0^\dagger\mathcal{X}_0\right]}{d^2-1}$$

equals one. Otherwise the *unitarity* $u(\mathcal{X}) \in [0,1]$ is strictly smaller. For instance, $u(\mathcal{D}) = 0$ for the depolarizing channel. This definition of the unitarity is equivalent to the one presented in Eq. (6), see [33, Proposition 1]. The argument outlined above succinctly summarises the main motivation for this figure of merit: it captures the coherence of a noise channel $\mathcal{X}$.

Equipped with this characterisation of the unitarity, we can now give the proof for the interpretation of the unitarity as the variance of the AGF with respect to a unitary 2-design.

Proof of Theorem 3. The unitarity $u(\mathcal{X})$ may be expressed as

$$\frac{\mathrm{Tr}\left[\mathcal{X}_0^\dagger\mathcal{X}_0\right]}{d^2-1} = \frac{\mathrm{Tr}\left[\left(\mathcal{X}_0 \oplus (d^2-1)x_1\right)^\dagger\mathcal{X}\right]}{d^2-1} - x_1^2. \tag{S.174}$$

Eq. (S.168) allows us to rewrite x_1 as an average over a unitary 1-design $\{\mathcal{U}_k\}_{k=1}^N$:

$$x_1 = \mathrm{Tr}\left[(\mathbb{O}\oplus 1)^\dagger\mathcal{X}\right] = \frac{1}{N}\sum_{k=1}^{N}\mathrm{Tr}\left[\mathcal{U}_k^\dagger\mathcal{X}\right] = \mathbb{E}\,\mathrm{Tr}\left[\mathcal{U}^\dagger\mathcal{X}\right]$$

Let us now assume that the set $\{\mathcal{U}_k\}_{k=1}^N$ is also a 2-design. Then, Eq. (S.159) implies

$$\frac{\left(\mathcal{X}_0 \oplus (d^2-1)x_1\right)^\dagger}{d^2-1} = \sum_{k=1}^{n} \mathcal{U}_k^\dagger \overline{\mathrm{Tr}\left[\mathcal{U}_k^\dagger \mathcal{X}\right]} = \mathbb{E}\, \mathcal{U}^\dagger \,\mathrm{Tr}\left[\mathcal{X}^\dagger \mathcal{U}\right]$$

Inserting both expressions into Eq. (S.174) yields

$$\begin{aligned} u(\mathcal{X}) &= \mathrm{Tr}\left[\mathcal{X}^\dagger \mathbb{E}\, \mathcal{U}\, \mathrm{Tr}\left[\mathcal{U}^\dagger \mathcal{X}\right]\right] - \left(\mathbb{E}\,\mathrm{Tr}\left[\mathcal{X}^\dagger \mathcal{U}\right]\right)^2 \\ &= \mathbb{E}\, \left|\mathrm{Tr}\left[\mathcal{X}^\dagger \mathcal{U}\right]\right|^2 - \left(\mathbb{E}\,\mathrm{Tr}\left[\mathcal{X}^\dagger \mathcal{U}\right]\right)^2 \\ &= \mathrm{Var}\left[\mathrm{Tr}\left[\mathcal{X}^\dagger \mathcal{U}\right]\right], \end{aligned}$$

where we have used linearity of the expectation value and the fact that the random variable $\mathrm{Tr}\left[\mathcal{X}^\dagger \mathcal{U}\right]$ is real-valued. Finally, we employ the relation between $\mathrm{Tr}\left[\mathcal{U}^\dagger \mathcal{X}\right]$ and $F_{\mathrm{avg}}(\mathcal{U}, \mathcal{X})$ presented in Eq. (S.161) to conclude

$$\begin{aligned} u(\mathcal{X}) &= \mathrm{Var}\left[\mathrm{Tr}\left[\mathcal{U}^\dagger \mathcal{X}\right]\right] \\ &= \mathrm{Var}\left[d(d+1)F_{\mathrm{avg}}(\mathcal{U},\mathcal{X}) - \mathrm{Tr}(\mathcal{X}(\mathrm{Id}))\right] \\ &= (d(d+1))^2\, \mathrm{Var}\left[F_{\mathrm{avg}}(\mathcal{U},\mathcal{X})\right], \end{aligned}$$

because variances are invariant under constant shifts and depend quadratically on scaling factors. This establishes Theorem 3. □

We conclude this section with a more speculative note regarding the possible applications for Theorem 3. A direct estimation procedure for the unitarity has been proposed in Ref. [33]. Inspired by randomised benchmarking, this procedure is robust towards SPAM errors, but has other drawbacks: Estimating the purity of outcome states directly is challenging, because the operator square function is not linear. Although Wallman et al. have found ways around this issue, their approaches are not yet completely satisfactory.

We propose an alternative approach based on Theorem 3. It might be conceivable that techniques like importance sampling could be employed to efficiently estimate this variance – and thus the unitarity – from "few" samples. The fourth moment bounds computed here could potentially serve as bounds on the "variance of this variance" and help control the convergence.

H. Numerical demonstrations

We emphasise that the main contributions of this work are of theoretical nature (we prove several Theorems). Nonetheless, we would also like to demonstrate the practical feasibility of our reconstruction procedure (3) and discuss some of its subtleties. The Matlab code for our numerical experiments can be found on GitHub [34].

Let $\mathcal{X}$ denote a unitary quantum channel. Given measurements f_i from Eq. (S.77) with Clifford unitaries C_i we approximately recover $\mathcal{X}$ using the semi-definite program (SDP) (S.79) with $q = 2$. In the numerical experiments we draw a three-qubit unitary channel $\mathcal{X}$ uniformly at random, the m Clifford unitaries for the measurements uniformly at random, and the noise $\epsilon \in \mathbb{R}^m$ uniformly from a sphere with radius η, i.e., $\|\epsilon\|_{\ell_2} = \eta$.

Then we solve the SDP using Matlab, CVX and SDPT3. The resulting average reconstruction error is plotted against the number of measurement settings m and the noise strength η in Figure 1 and Figure S.1 (left), respectively. As a comparison we run simulations for Haar random unitary measurements, see Figure S.1 (right). We find that the measurements based on random Clifford unitaries perform equally well as measurements based on Haar random unitaries. This is in agreement with a similar observation made for the noiseless case by two of the authors in Ref. [11].

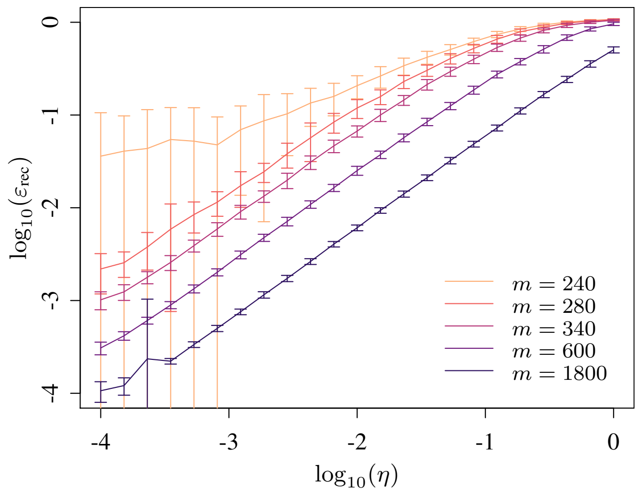
We observed that sometimes the SDP solver cannot find a solution. We also tested the use of Mosek instead of SDPT3. We find that the Mosek solver is faster but has more problems finding the correct solution. For the cases where the SDP solver does not exit with status "solved" we relax the machine precision on the equality constraints in the SDP (S.79) and change the measurement noise by a machine precision amount. More explicitly, for an integer $j \geq 0$ we try to solve

$$\begin{aligned} \underset{Z}{\text{minimise}} \quad & \|\mathcal{A}(Z) - f\|_{\ell_2} \\ \text{subject to} \quad & Z \geq 0, \\ & \left\|\mathrm{Tr}_1(Z) - \frac{\mathbb{1}}{d}\right\|_2 \leq 10^j \,\mathrm{eps}, \\ & \left\|\mathrm{Tr}_2(Z) - \frac{\mathbb{1}}{d}\right\|_2 \leq 10^j \,\mathrm{eps} \end{aligned} \tag{S.175}$$

where eps denotes the machine precision and Tr_1 and Tr_2 the partial traces on $\mathrm{L}(\mathbb{C}^d \otimes \mathbb{C}^d)$. We successively try to solve these SDPs for $j = 0, 1, 2, \ldots, 6$. Moreover, we change the measurement noise ϵ to $\epsilon' + \zeta$ in each of these trials, where each $\zeta_i = \mathrm{eps} \cdot g_i$ with $g_i \sim \mathcal{N}(0,1)$ is an independent normally distributed random number. For the Clifford type measurement (Figures 1 and S.1 left) a total of 24 400 channels were reconstructed and j was increased 1 865 many times in total. For the Haar random measurement unitaries (Figure S.1 left) a total of 12 900 channels were reconstructed and j was increased 950 times. So, we observed that with a probability of $\sim 7.5\%$ the SDP solver cannot solve the given SDP with machine precision constraints.

Some of the error bars in the plots in Figures 1 and S.1 might seem quite large, which we would like to comment on. Note that in compressed sensing it is typical to have a phase-transition from having no recovery for too small numbers of measurements m to having a recovery with very high probability once m exceeds a certain threshold. This phase transition region becomes smeared out if the noise strength $\|\epsilon\|_{\ell_2}$ is increased. For those m in the phase transition region the reconstruction errors are expected to fluctuate a lot, which we observe in the plots.

The slope of the linear part of plots $\varepsilon_{\mathrm{rec}}(m)$ in Figure 1 is roughly $\delta\varepsilon_{\mathrm{rec}}(m)/\delta m \approx -1.3$. This means that the reconstruction error scales like $\varepsilon_{\mathrm{rec}}(m) \sim m^{-1.3}$, which is better than Theorem 2 suggests. The reason for this discrepancy is that the theorem also bounds systematic errors and even adversarial noise ϵ whereas in the numerics we have drawn ϵ_i uniformly from a sphere, i.e., ϵ_i are i.i.d. up to a rescaling.



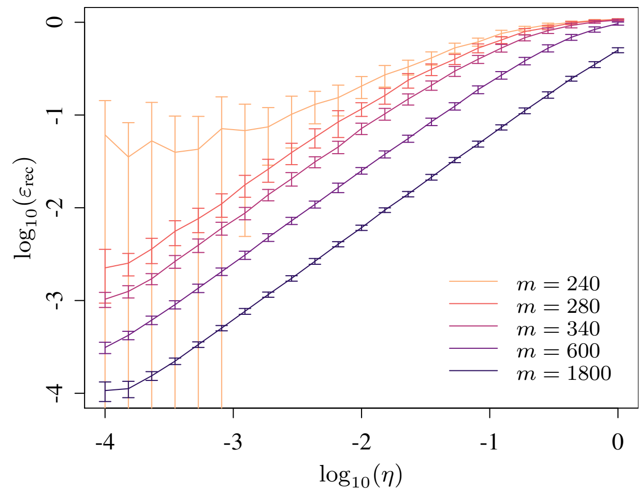


Figure S.1. Comparison of the reconstruction (3) from AGFs (2) with random Clifford unitaries (left) and Haar random unitaries (right). The plots show the dependence of the observed average reconstruction error $\varepsilon_{\text{rec}} := \left\| \mathcal{Z}^\sharp - \mathcal{X} \right\|$, on the noise strength $\eta := \|\epsilon\|_{\ell_2}$ for 3 qubits and different numbers of AGFs m. The error bars denote the observed standard deviation. The averages are taken over 100 samples of random i.i.d. measurements and channels (non-uniform). The Matlab code and data used to create these plots can be found on GitHub [34].

[1] M. Horodecki, P. Horodecki, and R. Horodecki, *General teleportation channel, singlet fraction, and quasidistillation*, Phys. Rev. A **60**, 1888 (1999).

[2] M. A. Nielsen, *A simple formula for the average gate fidelity of a quantum dynamical operation*, Phys. Lett. A **303**, 249 (2002), quant-ph/0205035.

[3] R. Kueng, D. M. Long, A. C. Doherty, and S. T. Flammia, *Comparing Experiments to the Fault-Tolerance Threshold*, Phys. Rev. Lett. **117**, 170502 (2016), arXiv:1510.05653 [quant-ph].

[4] B. Collins, *Moments and cumulants of polynomial random variables on unitary groups, the Itzykson-Zuber integral, and free probability*, Int. Math. Res. Not. **2003**, 953 (2003), math-ph/0205010.

[5] B. Collins and P. Sniady, *Representations of Lie groups and random matrices*, Trans. Amer. Math. Soc. **361**, 3269 (2009).

[6] R. Goodman and N. R. Wallach, *Symmetry, representations, and invariants* (Springer, Berlin, 2009).

[7] W. Fulton and J. Harris, *Representation theory*, Vol. 129 (Springer Science & Business Media, 1991).

[8] E. P. Wigner, *Group theory and its application to the quantum mechanics of atomic spectra* (Academic Press, London, 1959).

[9] This way of stating the result of Ref. [5] was brought to our attention by study notes of K. Audenaert.

[10] H. Zhu, R. Kueng, M. Grassl, and D. Gross, *The Clifford group fails gracefully to be a unitary 4-design*, arXiv:1609.08172 [quant-ph].

[11] S. Kimmel and Y. K. Liu, Phase retrieval using unitary 2-designs, in *2017 International Conference on Sampling Theory and Applications (SampTA)* (2017) pp. 345–349, arXiv:1510.08887 [quant-ph].

[12] R. Kueng, H. Rauhut, and U. Terstiege, *Low rank matrix recovery from rank one measurements*, Appl. Comp. Harm. Anal. (2015), arXiv:1410.6913 [cs.IT].

[13] M. Kabanava, R. Kueng, H. Rauhut, and U. Terstiege, *Stable low-rank matrix recovery via null space properties*, Information and Inference: A Journal of the IMA **5**, 405 (2016), arXiv:1507.07184 [cs.IT].

[14] S. Foucart and H. Rauhut, *A mathematical introduction to compressive sensing* (Springer, 2013).

[15] M. Kliesch, R. Kueng, J. Eisert, and D. Gross, *Guaranteed recovery of quantum processes from few measurements*, arXiv:1701.03135 [quant-ph].

[16] J. A. Tropp, *Convex recovery of a structured signal from independent random linear measurements*, in *Sampling Theory, a Renaissance*, edited by E. G. Pfander (Springer, 2015) pp. 67–101, arXiv:1405.1102 [cs.IT].

[17] S. Mendelson, *Learning without concentration*, J. ACM **62**, 21:1 (2015), arXiv:1401.0304 [cs.LG].

[18] V. Koltchinskii and S. Mendelson, *Bounding the smallest singular value of a random matrix without concentration*, International Mathematics Research Notices **2015**, rnv096 (2015), arXiv:1312.3580 [math.PR].

[19] R. Vershynin, *Introduction to the non-asymptotic analysis of random matrices*, in *Compressed Sensing: Theory and Applications* (Cambridge University Press, 2012) pp. 210–268, arXiv:1011.3027 [math.PR].

[20] J. A. Tropp, *User friendly tools for random matrices. An introduction.* Preprint (2012).

[21] H. Zhu, *Multiqubit Clifford groups are unitary 3-designs*, Phys. Rev. A **96**, 062336 (2017), arXiv:1510.02619 [quant-ph].

[22] Z. Webb, *The Clifford Group Forms a Unitary 3-design*, Quantum Info. Comput. **16**, 1379 (2016), arXiv:1510.02769 [quant-ph].

[23] S. T. Flammia and Y.-K. Liu, *Direct fidelity estimation from few Pauli measurements*, Phys. Rev. Lett. **106**, 230501 (2011), arXiv:1104.4695 [quant-ph].

[24] S. T. Flammia, D. Gross, Y.-K. Liu, and J. Eisert, *Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators*, New J. Phys. **14**, 095022 (2012), arXiv:1205.2300 [quant-ph].

[25] M. P. da Silva, O. Landon-Cardinal, and D. Poulin, *Practical characterization of quantum devices without tomography*, Phys. Rev. Lett. **107**, 210404 (2011), arXiv:1104.3835 [quant-ph].

[26] A. R. Calderbank, E. M. Rains, P. W. Shor, and N. J. A. Sloane, *Quantum error correction via codes over GF(4)*, IEEE Trans. Inf. Theory **44**, 1369 (1998), arXiv:quant-ph/9608006.

[27] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, *Sample-Optimal Tomography of Quantum States*, IEEE Trans. Inf. Theory **63**, 5628 (2017), arXiv:1508.01797 [quant-ph].

[28] T. M. Cover and J. A. Thomas, *Elements of information theory* (John Wiley & Sons, 2012).

[29] P. Diaconis and M. Shahshahani, *On the Eigenvalues of Random Matrices*, Journal of Applied Probability **31**, 49 (1994).

[30] A. J. Scott, *Optimizing quantum process tomography with unitary 2-designs*, J. Phys. A **41**, 055308 (2008), arXiv:0711.1017 [quant-ph].

[31] More explicitly, for $X \in H_d$ we can calculate $\int_{U(d)} \mathcal{U}(X)\, d\mu(U) = E^1_{\Delta_{U(d)}}(X) = \mathbb{O} \oplus 1$ using Theorem S.2.

[32] D. Gross, K. Audenaert, and J. Eisert, *Evenly distributed unitaries: on the structure of unitary designs*, J. Math. Phys. **48**, 052104, quant-ph/0611002.

[33] J. Wallman, C. Granade, R. Harper, and S. T. Flammia, *Estimating the coherence of noise*, New J. Phys. **17**, 113020 (2015), arXiv:1503.07865 [quant-ph].

[34] I. Roth, R. Kueng, S. Kimmel, Y.-K. Liu, D. Gross, J. Eisert, and M. Kliesch, *Quantum process tomography with average gate fidelities*, GitHub repository `https://github.com/MartKl/Quantum_process_tomography_with_average_gate_fidelities` (2017).